Multi-dimensional feature and data mining

Georgiou, T.

APPENDICIES

A Clifford Convolution gradients calculations

Bellow are all the calculations of all the equations used for the back-propagation algorithm. For all formulas, for the sake of simplicity we use the following notation:

$$\begin{aligned} &p_w : \text{weight position, } (i, j, c_{in}, c_{out}) \\ &p_o : \text{output position, } (i', j', c_{out}) \\ &p_{in} : \text{input position, } (i'', j'', c_{in}) \end{aligned} \tag{1}$$

During the backward pass, every arrow in Figure 5.2 should return the respective derivatives. We start the computations from step C_5 and work our way to step C_1.

C_5:
The final output $\vec{O}^l_{p_o}$ is given by:

$$\vec{O}^l_{p_o} = (O^l_{0,p_o}, O^l_{1,p_o}) = (O^l_{p_o} \cdot \cos(\phi_{p_o}), O^l_{p_o} \cdot \sin(\phi_{p_o})) \tag{2}$$

Following the chain rule:

$$\begin{aligned} &\frac{\partial E}{\partial O^l_{p_o}} = \frac{\partial E}{\partial O^l_{0,p_o}} \frac{\partial O^l_{0,p_o}}{\partial O^l_{p_o}} + \frac{\partial E}{\partial O^l_{1,p_o}} \frac{\partial O^l_{1,p_o}}{\partial O^l_{p_o}} \Rightarrow \\ &\frac{\partial E}{\partial O^l_{p_o}} = \delta^{l,C_5}_{0,p_o} \cdot \cos(\phi_{p_o}) + \delta^{l,C_5}_{1,p_o} \cdot \sin(\phi_{p_o}) \end{aligned} \tag{3}$$

$$\begin{aligned}\frac{\partial E}{\partial \phi^l_{p_o}} &= \frac{\partial E}{\partial O^l_{0,p_o}}\frac{\partial O^l_{0,p_o}}{\partial \phi^l_{p_o}} + \frac{\partial E}{\partial O^l_{1,p_o}}\frac{\partial O^l_{1,p_o}}{\partial \phi^l_{p_o}} \Rightarrow \\ \frac{\partial E}{\partial \phi^l_{p_o}} &= -\delta^{l,C_5}_{0,p_o} \cdot O^l_{0,p_o} \cdot \sin(\phi_{p_o}) + \delta^{l,C_5}_{1,p_o} \cdot O^l_{1,p_o} \cdot \cos(\phi_{p_o})\end{aligned} \tag{4}$$

C_4:

$$O^l_{p_o} = \sum_i \sum_j \sum_{c_{in}} \vec{W}^l_{\phi\ p_o p_w} \cdot \vec{O}^{l-1}_{p_{in}} = \sum_i \sum_j \sum_{c_{in}} \sum_k W^l_{\phi\ p_o p_w, k} \cdot O^{l-1}_{p_{in},k} \tag{5}$$

where k indexes the x and y coordinates of the vectors.

$$\frac{\partial E}{\partial W^l_{\phi\ p_o p_w, k}} = \frac{\partial E}{\partial O^l_{p_o}} \frac{\partial O^l_{p_o}}{\partial W^l_{\phi\ p_o p_w, k}} \tag{6}$$

Since for every output pixel we calculate a different angle ($\phi^l_{p_o}$), there is a different set of weights associated with every output pixel, $\vec{W}^l_{\phi\ p_o}$.

$$\frac{\partial E}{\partial W^l_{\phi\ p_o p_w, k}} = \frac{\partial E}{\partial O^l_{p_o}} \frac{\partial O^l_{p_o}}{\partial W^l_{\phi\ p_o p_w, k}} = \sum_{p_o \in P_\phi} \delta^{l,C_4}_{p_o} \cdot O^{l-1}_{p_{in},k} \tag{7}$$

$$\frac{\partial E}{\partial O^{l-1}_{p_{in},k}} = \sum_i \sum_j \sum_{c_{out}} \frac{\partial E}{\partial O^l_{p_o}} \frac{\partial O^l_{p_o}}{\partial O^{l-1}_{p_{in},k}} = \sum_i \sum_j \sum_{c_{out}} \delta^{l4}_{p_o} \cdot W^l_{\phi\ p_o p_w, k} \tag{8}$$

where the ϕ in $W^l_{\phi\ p_o p_w, k}$ refers to the predefined angles ($\in [0, 2\pi)$) used for the specific output position.

C_3:

$$\vec{W}^l_{\phi\ p_o} = rotation(\dot{\vec{W}}^l_{p_o}, \phi^l_{p_o}) \tag{9}$$

As mentioned above, there is a seperate set of weights for every output pixel. The gradients from all sets of weights are calculated and then added to the orginal weight vector. For simplicity the index of p_o on the weight vectors is omited. From equation 9 we see that there are two sets of gradients need to be computed, $\frac{\partial E}{\partial \dot{\vec{W}}^l}$ and $\frac{\partial E}{\partial \phi^l_{p_o}}$. Since $\vec{W}^l_\phi$ are the rotated $\dot{\vec{W}}^l$, we set:

$$\frac{\partial E}{\partial \dot{\vec{W}}^l} = rotation(\frac{\partial E}{\partial \vec{W}^l_\phi}, -\phi^l_{p_o}) \tag{10}$$

For the second set we have:

$$\frac{\partial E}{\partial \phi^l_{p_o}} = \sum_i \sum_j \sum_{c_{in}} \frac{\partial E}{\partial \vec{W}^l_{\phi\ p_w}} \frac{\partial \vec{W}^l_{\phi\ p_w}}{\partial \phi^l_{p_o}} = \sum_i \sum_j \sum_{c_{in}} \delta^{l,C_3}_{\vec{W}_{\phi\ p_w}} \cdot \frac{\partial \vec{W}^l_{\phi\ p_w}}{\partial \phi^l_{p_o}} \tag{11}$$

We have two options for calculating $\frac{\partial \vec{W}^l_{\phi\ p_w}}{\partial \phi^l_{p_o}}$. The first is to differentiate the bilinear interpolation (at least at the points that it is differentiable), or use the precalculated rotated weights. Let θ be the quantized calculated angle ϕ. Then:

$$\frac{\partial \vec{W}^l_{\theta\ p_w}}{\partial \phi^l_{p_o}} = \frac{\vec{W}^l_{\theta+1\ p_w} - \vec{W}^l_{\theta-1\ p_w}}{2\frac{2\pi}{B}} \tag{12}$$

where B is the number of predefined orientations. Have in mind that the rotation above ($\vec{W}^l_{\theta+1\ p_w}$) considers only plane rotation after acquiring the in-place vector rotation $\vec{\dot{W}}_\phi$.

C_2:

On all equations related to C_2, like C_3, the indexes p_o on weight vectors are omitted.

$$\vec{\dot{W}}^l_{\phi\ p_w} = \vec{W}^l_{p_w} \cdot \begin{pmatrix} \cos\phi^l_{p_o} & \sin\phi^l_{p_o} \\ -\sin\phi^l_{p_o} & \cos\phi^l_{p_o} \end{pmatrix} \Leftrightarrow \begin{cases} \vec{\dot{W}}^l_{\phi\ 0,p_w} = \vec{W}^l_{0,p_w}\cos\phi^l_{p_o} - \vec{W}^l_{1,p_w}\sin\phi^l_{p_o} \\ \vec{\dot{W}}^l_{\phi\ 1,p_w} = \vec{W}^l_{0,p_w}\sin\phi^l_{p_o} + \vec{W}^l_{1,p_w}\cos\phi^l_{p_o} \end{cases} \tag{13}$$

As with C_3, there are two set of gradients to be calculated, specifically $\frac{\partial E}{\partial \vec{W}^l_{p_w}}$ and $\frac{\partial E}{\partial \phi^l_{p_o}}$ where the first represents two sets, one for each direction of the vectors in $\vec{W}^l$.

$$\frac{\partial E}{\partial \vec{W}^l_{p_w}} = (\frac{\partial E}{\partial W^l_{0,p_w}}, \frac{\partial E}{\partial W^l_{1,p_w}}) \tag{14}$$

For the two components we get:

$$\begin{aligned} \frac{\partial E}{\partial W^l_{0,p_w}} &= (\frac{\partial E}{\partial \dot{W}^l_{\phi\ 0,p_w}} \frac{\partial \dot{W}^l_{\phi\ 0,p_w}}{\partial W^l_{0,p_w}} + \frac{\partial E}{\partial \dot{W}^l_{\phi\ 1,p_w}} \frac{\partial \dot{W}^l_{\phi\ 1,p_w}}{\partial W^l_{0,p_w}}) \\ &= (\frac{\partial E}{\partial \dot{W}^l_{\phi\ 0,p_w}} \cos\phi^l_{p_o} + \frac{\partial E}{\partial \dot{W}^l_{\phi\ 1,p_w}} \sin\phi^l_{p_o}) \\ \frac{\partial E}{\partial W^l_{1,p_w}} &= (\frac{\partial E}{\partial \dot{W}^l_{\phi\ 0,p_w}} \frac{\partial \dot{W}^l_{\phi\ 0,p_w}}{\partial W^l_{1,p_w}} + \frac{\partial E}{\partial \dot{W}^l_{\phi\ 1,p_w}} \frac{\partial \dot{W}^l_{\phi\ 1,p_w}}{\partial W^l_{1,p_w}}) \\ &= (-\frac{\partial E}{\partial \dot{W}^l_{\phi\ 0,p_w}} \sin\phi^l_{p_o} + \frac{\partial E}{\partial \dot{W}^l_{\phi\ 1,p_w}} \cos\phi^l_{p_o}) \end{aligned} \tag{15}$$

$$(14),(15) \rightarrow \frac{\partial E}{\partial \vec{W}^l_{p_w}} = \frac{\partial E}{\partial \dot{\vec{W}}^l_{\phi\ p_w}} \cdot \begin{pmatrix} \cos(-\phi^l_{p_o}) & \sin(-\phi^l_{p_o}) \\ -\sin(-\phi^l_{p_o}) & \cos(-\phi^l_{p_o}) \end{pmatrix}$$
$$= \vec{\delta}^{l,C_2}_{\phi\ p_w} \cdot \begin{pmatrix} \cos(-\phi^l_{p_o}) & \sin(-\phi^l_{p_o}) \\ -\sin(-\phi^l_{p_o}) & \cos(-\phi^l_{p_o}) \end{pmatrix} \tag{16}$$

$$\frac{\partial E}{\partial \phi^l_{p_o}} = \sum_i \sum_j \sum_{c_{in}} \left(\frac{\partial E}{\partial \dot{W}^l_{0,p_w}} \frac{\partial \dot{W}^l_{0,p_w}}{\partial \phi^l_{p_o}} + \frac{\partial E}{\partial \dot{W}^l_{1,p_w}} \frac{\partial \dot{W}^l_{1,p_w}}{\partial \phi^l_{p_o}} \right) \tag{17}$$

$$\frac{\partial \dot{W}^l_{0,p_w}}{\partial \phi^l_{p_o}} = -W^l_{0,p_w} \sin\phi^l_{p_o} - W^l_{1,p_w} \cos\phi^l_{p_o} = -\dot{W}^l_{1,p_w}$$
$$\frac{\partial \dot{W}^l_{1,p_w}}{\partial \phi^l_{p_o}} = W^l_{0,p_w} \cos\phi^l_{p_o} - W^l_{1,p_w} \sin\phi^l_{p_o} = \dot{W}^l_{0,p_w} \tag{18}$$

$$(17),(18) \rightarrow \frac{\partial E}{\partial \phi^l_{p_o}} = \sum_i \sum_j \sum_{c_{in}} \left(\frac{\partial E}{\partial \dot{W}^l_{0,p_w}} (-\dot{W}^l_{1,p_w}) + \frac{\partial E}{\partial \dot{W}^l_{1,p_w}} \dot{W}^l_{0,p_w} \right)$$
$$= \sum_i \sum_j \sum_{c_{in}} \left(\delta^{l,C_2}_{0,p_w} (-\dot{W}^l_{1,p_w}) + \delta^{l,C_2}_{1,p_w} \dot{W}^l_{0,p_w} \right) \tag{19}$$

In our implementation C_2 and C_3 are considered as one operation. Moreover, we keep in memory the rotated weights $\vec{W}_\phi$ and not $\dot{\vec{W}}_\phi$. Fortunately, we can approximate the gradients as of the separate operations as following:

$$\vec{W}^l_\phi = vector\_field\_rotation(\vec{W}^l, \phi^l_{p_o}) \tag{20}$$

Similarly with C_3:

$$\frac{\partial E}{\partial \vec{W}^l} = vector\_field\_rotation\left(\frac{\partial E}{\partial \vec{W}^l_\phi}, -\phi^l_{p_o}\right) \tag{21}$$

$$\frac{\partial \vec{W}^l_{\theta\ p_w}}{\partial \phi^l_{p_o}} = \frac{\vec{W}^l_{\theta+1\ p_w} - \vec{W}^l_{\theta-1\ p_w}}{2\frac{2\pi}{B}} \tag{22}$$

Unlike C_3, here the rotated $\vec{W}^l_{\theta+1\ p_w}$ are the complete vector field rotation with angle $\theta + 1$ from the original $\vec{W}$.

C_1:
Let:

$$tan^l_{p_o} = \frac{conv_0^{l_{p_o}}}{conv_2^{l_{p_o}}} \tag{23}$$

then:

$$(5.2), (5.1) \rightarrow \phi^l_{p_o} = \arctan\left(\frac{conv_2}{conv_0}\right) = \arctan\left(tan^l_{p_o}\right) \tag{24}$$

$$\frac{\partial E}{\partial tan^l_{p_o}} = \frac{\partial E}{\partial \phi^l_{p_o}} \frac{\partial \phi^l_{p_o}}{\partial tan^l_{p_o}} = \frac{\partial E}{\partial \phi^l_{p_o}} \frac{1}{1 + (tan^l_{p_o})^2} \tag{25}$$

$$\begin{gathered} \frac{\partial E}{\partial conv_0} = \frac{\partial E}{\partial tan^l_{p_o}} \frac{\partial tan^l_{p_o}}{\partial conv_0} = \frac{\partial E}{\partial tan^l_{p_o}} \left(-\frac{conv_2}{conv_0^2}\right) \overset{25}{\Rightarrow} \\ \frac{\partial E}{\partial conv_0} = \frac{\partial E}{\partial \phi^l_{p_o}} \frac{1}{1 + (tan^l_{p_o})^2} \left(-\frac{conv_2}{conv_0^2}\right) = -\frac{\partial E}{\partial \phi^l_{p_o}} \frac{conv_2}{conv_0^2 + conv_2^2} \end{gathered} \tag{26}$$

$$\begin{gathered} \frac{\partial E}{\partial conv_2} = \frac{\partial E}{\partial tan^l_{p_o}} \frac{\partial tan^l_{p_o}}{\partial conv_2} = \frac{\partial E}{\partial tan^l_{p_o}} \frac{1}{conv_0} \overset{25}{\Rightarrow} \\ \frac{\partial E}{\partial conv_2} = \frac{\partial E}{\partial \phi^l_{p_o}} \frac{1}{1 + (tan^l_{p_o})^2} \frac{1}{conv_0} = \frac{\partial E}{\partial \phi^l_{p_o}} \frac{conv_0}{conv_0^2 + conv_2^2} \end{gathered} \tag{27}$$

From Equation 5.2 we see that $conv_0$ is the conventional convolutional operation, meaning that the derivatives are the standard derivatives used in all CNN works. For $conv_2$ we have:

$$\begin{aligned} \frac{\partial E}{\partial W^l_{0,p_w}} &= \sum_{i'} \sum_{j'} \frac{\partial E}{\partial conv_2} \frac{\partial conv_2}{\partial W^l_{0,p_w}} = \sum_{i'} \sum_{j'} \frac{\partial E}{\partial conv_2} O^{l-1}_{1,p_{in}} \\ \frac{\partial E}{\partial W^l_{1,p_w}} &= \sum_{i'} \sum_{j'} \frac{\partial E}{\partial conv_2} \frac{\partial conv_2}{\partial W^l_{1,p_w}} = \sum_{i'} \sum_{j'} \frac{\partial E}{\partial conv_2} (-O^{l-1}_{0,p_{in}}) \end{aligned} \tag{28}$$

Similarly:

$$\begin{aligned} \frac{\partial E}{\partial O^{l-1}_{0,p_{in}}} &= \sum_{i} \sum_{j} \frac{\partial E}{\partial conv_2} \frac{\partial conv_2}{\partial O^{l-1}_{0,p_{in}}} = \sum_{i} \sum_{j} \frac{\partial E}{\partial conv_2} (-W^l_{1,p_w}) \\ \frac{\partial E}{\partial O^{l-1}_{1,p_{in}}} &= \sum_{i} \sum_{j} \frac{\partial E}{\partial conv_2} \frac{\partial conv_2}{\partial O^{l-1}_{1,p_{in}}} = \sum_{i} \sum_{j} \frac{\partial E}{\partial conv_2} W^l_{0,p_w} \end{aligned} \tag{29}$$

For each output pixel a separate weight vector was calculated and thus different gradients as well, i.e., $\left(\frac{\partial E}{\partial \vec{W}}\right)_{p_o}$. The final result is given by adding the $\left(\frac{\partial E}{\partial \vec{W}}\right)_{p_o}$ for all p_o.

B Table of abbreviations

Abbreviation	Explanation
2D	two dimensions/dimensional
3D	three dimensions/dimensional
3DBRIEF	3D BRIEF
3DLBP	3D LBP
3DORB	3D ORB
3DSC	3D SC
4D	four dimensions/dimensional
Adam	adaptive moment estimation
AE	auto-encoder
AGAST	adaptive and generic accelerated segment test
AlexNet	Alex Network
AMT	Amazon mechanical turk
ANN	artificial neural network
APC	Amazon picking challenge
API	application programming interface
avacc	meanIU
B3DO	Berkley 3D Objects
BN	batch normalization
BoF	bag of features
BoW	bag of words
BPTT	back propagation through time
BRAND	binary robust appearance and normal descriptor
BRIEF	binary robust independent elementary features
BRISK	binary robust invariant scale keypoint
BRoPH	binary rotational projection histogram
C3D	convolutional 3D
CAD	computer-aided design
CAE	convolutional AE

Abbreviation	Explanation
CBCT	cone beam computed tomography
cc	Clifford convolution
CFD	computational fluid dynamics
CFN	convolutional fusion network
Charades-STA	Charades sentence temporal annotations
CHMM	coupled HMM
CIFAR	Canadian institute for advanced research
CL	convolutional layer
clacc	classification accuracy
CNN	convolutional neural network
COCO	common objects in context
convGRBM	convolutional GRBM
CPU	central processing unit
CRF	conditional random field
CT	computerized tomography
DAE	denoising AE
DB	database
DBM	deep Boltzmann machines
DBN	deep belief network
D-CNN	deep CNN
DE	dense sampling
DEM	deep energy model
DenseNet	dense network
DiDeMo	distinct describable moments
DL	deep learning
DNN	deep neural network
DoG	difference of Gaussians
DoF	degrees of freedom
DS	direction specific
DSN	deeply supervised nets
DSTIP	depth STIP
ED	elevation descriptor
ELU	exponential linear unit
EMK	efficient match kernel
EVD	eigenvalue decomposition
FAST	features from accelerated segment test
FC	fully connected

Abbreviation	Explanation
FCN	fully convolutional networks
FCVID	Fudan-Columbia video dataset
FMS	full modality specific
FPFH	fast PFH
FREAK	fast retina keypoint
fus-CNN	fusion CNN
fwavacc	frequency weighted average accuracy
GAH	geometric attribute histograms
GAN	generative adversarial network
GFU	gated fusion unit
GNN	graph neural network
GPU	graphics processing unit
GRBM	gated RBM
GRU	gated recurrent unit
GT	ground truth
HAR	human action recognition
Harris3D	Harris 3D
HBN	half layers batch normalized
HCRF	hidden CRF
HHA	horizontal disparity, height above ground, angle the pixels local surface normal makes with the inferred gravity direction
HKDE	hierarchical KDE
HKS	heat kernel signature
HMC	hidden Markov chain
HMDB51	human motion database
HMM	hidden Markov model
HMP	hierarchical matching pursuit
HOF	histogram of flow
HOG	histogram of oriented gradients
HON	histogram of surface normals
HON4D	HON 4D
HOPC	histogram of principal components
HSMM	hidden semi-Markov model
I3D	inflated 3D CNN
IDT	improved dense trajectories
IP	interest point
KDE	kernel descriptor

Abbreviation	Explanation
kd-tree	k dimensional tree
KITTI	? (not mentioned in the work that proposes it [98])
KLT	Kanade Lucas-Tomasi
k-NN	k nearest neighbors
kSVM	kernel SVM
KTH	Royal institute of technology, Stockholm
LBP	local binary pattern
LeNet	LeCun network
LFD	light field descriptor
LFSH	local feature statistics histogram
LiDAR	light detection and ranging
LINE	linearizing the memory
LINEMOD	multimodal LINE
linSVM	linear SVM
LN	locally connected
LRCN	long-term recurrent CNN
LReLU	leaky ReLU
LRF	local reference frame
LSP	local surface patch
LSTM	long short-term memory node
LSTM-CF	LSTM context fusion
LTC	long temporal convolutional network
LTP	local trinary pattern
MAE	mean absolute error
MD	multiple dictionary
meanIU	mean intersection over union
MK-MMD	multiple kernel maximum mean discrepancy
MLP	multi-layer perceptron
MMF	multi modal feature fusion
MNIST	modified national institute of standards and technology
MO-AniProbing	multi orientation anisotropic probing
mp	max pooling
MR	magnetic resonance
MRF	Markov random field
MRI	magnetic resonance imaging
MVCNN	multi view CNN
MVD	multi-view depth

Abbreviation	Explanation
NaN	not a number
NBN	no BN
NiN	network in network
nl	norm loss
NN	nearest neighbor
NNDR	nearest neighbor distance ratio
NYU	New York University
NYUv2	NYU version 2
OGH	oriented gradient histograms
OLM	orthogonal linear module
ONI	orthogonalization using Newton's iteration
op	orientation pooling
ORB	oriented FAST and rotated BRIEF
ORION	orientation boosted voxel net
ORN	orientation response network
PA-LSTM	part-aware LSTM
PBWN	projection based weight normalization
PCA	principal component analysis
PELU	parametric ELU
PFH	point feature histogram
pixacc	pixel accuracy
PPF	point pair feature
PReLU	parametric ReLU
PSB	Princeton shape benchmark
PSG	polygonal surface geometry
RA	reference angle
RANSAC	random sample consensus
RAS	Reynolds-averaged simulation
RBM	restricted Boltzmann machine
R-CNN	regions with CNN features
RDF	randomized decision forest
RDF-Net	RGB-D fusion network
ReLU	rectified linear unit
ResBlock	residual block
ResNet	residual network
RF	random forest
RFB	residual fusion block

Abbreviation	Explanation
RGB	Red-Green-Blue
RGB-D	Red-Green-Blue-Depth
RI-LBC	rotation invariant local binary convolution
RMSE	root mean square error
RNN	recurrent neural network
Rohr3D	Karl Rohr 3D
RoSP	rotational projection statistics
RotEqNet	rotation equivariant vector field network
RQ	research question
RSM	rotational silhouette map
SC	shape context
SD	single dictionary
SDH	spatial distribution histograms
SF	sparse fusion
SFCNN	steerable filter CNN
SfM	structure from motion
SGD	stochastic gradient descent
SHOT	signature of histograms of orientation
SHREC	shape retrieval contest
SI	spin image
SIFT	scale invariant feature transform
SI-HKS	SI HKS
SISI	scale invariant SI
SLAM	simultaneous localization and mapping
SP	superpixel
SPN	scalar field processing network
SRIP	spectral restricted isometry property
SSCD	spatial structure circular descriptor
SSD	sum of squared differences
SSMA	self-supervised model adaptation
SSVM	structural SVM
std	standard deviation
STIP	spatio-temporal interest point
ST-LSTM	spatio-temporal LSTM
STN	spatial transform networks
SUN	scene understanding
SUN-CG	? (not mentioned in the work that proposes it [346])

Abbreviation	Explanation
SURF	speeded up robust features
SVD	singular value decomposition
SVM	support vector machine
SYNTHIA	synthetic collection of imagery and annotations
TACoS	textually annotated cooking scenes
TDD	trajectory pooled deep convolutional descriptors
THRIFT	? (not mentioned in the work that proposes it [90])
TI	transformation invariant
TOLDI	triple orthogonal local depth images
Tri-SI	Tri-Spin-Image
UCF	university of central Florida
UMAM	unified model of appearance and motion
US	ultrasound
VC	velocity coherent
V-FAST	video FAST
VFT	vector field topology
VGG	? (not mentioned in the work that proposes it [338])
VPN	vector processing network
VRN	Voxeption ResNet
wd	weight decay
WKS	wave kernel signature
WN	weight normalization
WRN	wide ResNet
YCB	Yale-CMU-Berkeley
YFCC100M	Yahoo Flickr creative commons 100 million

Bibliography

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th symposium on operating systems design and implementation (OSDI)*, pages 265–283, 2016.

[2] P-A Absil and Kyle A Gallivan. Joint diagonalization on the oblique manifold for independent component analysis. In *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 5, pages V–V, 2006.

[3] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[4] Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*, 2016.

[5] Forest Agostinelli, Matthew Hoffman, Peter Sadowski, and Pierre Baldi. Learning activation functions to improve deep neural networks. *arXiv preprint arXiv:1412.6830*, 2014.

[6] Motilal Agrawal, Kurt Konolige, and Morten Rufus Blas. Censure: Center surround extremas for realtime feature detection and matching. In *Proceedings of the European conference on computer vision (ECCV)*, pages 102–115. Springer, 2008.

[7] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak: Fast retina keypoint. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 510–517. IEEE, 2012.

[8] Luís A Alexandre. 3d object recognition using convolutional neural networks with transfer learning between input channels. In *Intelligent Autonomous Systems 13*, pages 889–898. Springer, 2016.

[9] Stéphane Allaire, John J Kim, Stephen L Breen, David A Jaffray, and Vladimir Pekar. Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 1–8. IEEE, 2008.

[10] Lisa Anne Hendricks, Oliver Wang, Eli Shechtman, Josef Sivic, Trevor Darrell, and Bryan Russell. Localizing moments in video with natural language. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 5803–5812. IEEE, 2017.

[11] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Proceedings of the IEEE international conference on computer vision workshops (ICCVW)*, pages 1626–1633. IEEE, 2011.

[12] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[13] Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Sequential deep learning for human action recognition. In *International workshop on human behavior understanding*, pages 29–39. Springer, 2011.

[14] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 39:2481–2495, 2017.

[15] Randall Balestriero et al. A spline theory of deep learning. In *International Conference on Machine Learning (ICML)*, pages 374–383, 2018.

[16] Vassileios Balntas, Andreas Doumanoglou, Caner Sahin, Juil Sock, Rigas Kouskouridas, and Tae-Kyun Kim. Pose guided rgbd feature learning for 3d object pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3856–3864. IEEE, 2017.

[17] Nitin Bansal, Xiaohan Chen, and Zhangyang Wang. Can we gain more from orthogonality regularizations in training deep networks? In *Advances in Neural Information Processing Systems*, pages 4261–4271, 2018.

[18] Mohammadamin Barekatain, Miquel Martí, Hsueh-Fu Shih, Samuel Murray, Kotaro Nakayama, Yutaka Matsuo, and Helmut Prendinger. Okutama-action: An aerial view video dataset for concurrent human action detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 28–35. IEEE, 2017.

[19] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 404–417. Springer, 2006.

[20] Paul R Beaudet. Rotationally invariant image operators. In *Proceedings of the 4th international joint conference on pattern recognition*, 1978.

[21] Jens Behley, Volker Steinhage, and Armin B Cremers. Laser-based segment classification using a mixture of bag-of-words. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 4195–4200. IEEE, 2013.

[22] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 24:509–522, 2002.

[23] Hakan Bilen, Basura Fernando, Efstratios Gavves, Andrea Vedaldi, and Stephen Gould. Dynamic image networks for action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3034–3042. IEEE, 2016.

[24] Michael J Black and Allan D Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International journal of computer vision (IJCV)*, 26:63–84, 1998.

[25] Liefeng Bo, Kevin Lai, Xiaofeng Ren, and Dieter Fox. Object recognition with hierarchical kernel descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1729–1736. IEEE, 2011.

[26] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Kernel descriptors for visual recognition. In *Advances in neural information processing systems 23*, pages 244–252. Curran Associates, Inc., 2010.

[27] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Depth kernel descriptors for object recognition. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 821–826. IEEE, 2011.

[28] Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Unsupervised feature learning for rgb-d based object recognition. In Jaydev P. Desai, Gregory Dudek, Oussama Khatib, and Vijay Kumar, editors, *Experimental Robotics*, pages 387–402. Springer, 2013.

[29] Aaron F. Bobick and James W. Davis. The recognition of human movement using temporal templates. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 23:257–267, 2001.

[30] Ujwal Bonde, Vijay Badrinarayanan, and Roberto Cipolla. Robust instance recognition in presence of occlusion and clutter. In *Proceedings of the European conference on computer vision (ECCV)*, pages 520–535. Springer, 2014.

[31] Hervé Bourlard and Yves Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59:291–294, 1988.

[32] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[33] Matteo Bregonzio, Shaogang Gong, and Tao Xiang. Recognising action as clouds of space-time interest points. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1948–1955. IEEE, 2009.

[34] Rasmus Bro, Evrim Acar, and Tamara G Kolda. Resolving the sign ambiguity in the singular value decomposition. *Journal of Chemometrics*, 22:135–140, 2008.

[35] Andrew Brock, Theodore Lim, JM Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. *arXiv preprint arXiv:1608.04236*, 2016.

[36] Alexander Bronstein, Michael Bronstein, and Maks Ovsjanikov. 3d features, surface descriptors, and object descriptors. *3D Imaging, Analysis, and Applications*, pages 1–27, 2010.

[37] Alexander M Bronstein, Michael M Bronstein, Leonidas J Guibas, and Maks Ovsjanikov. Shape google: Geometric words and expressions for invariant shape retrieval. *ACM Transactions on Graphics (TOG)*, 30:1, 2011.

[38] Michael M Bronstein and Iasonas Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1704–1711. IEEE, 2010.

[39] Fabian Caba Heilbron, Victor Escorcia, Bernard Ghanem, and Juan Carlos Niebles. Activitynet: A large-scale video benchmark for human activity understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 961–970. IEEE, 2015.

[40] Berk Calli, Arjun Singh, Aaron Walsman, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. The ycb object and model set: Towards common benchmarks for manipulation research. In *International conference on advanced robotics (ICAR)*, pages 510–517. IEEE, 2015.

[41] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 778–792, 2010.

[42] Liangliang Cao, Zicheng Liu, and Thomas S Huang. Cross-dataset action detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1998–2005. IEEE, 2010.

[43] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 4724–4733. IEEE, 2017.

[44] Bhaskar Chakraborty, Michael B Holte, Thomas B Moeslund, and Jordi Gonzàlez. Selective spatio-temporal interest points. *Computer vision and image understanding (CVIU)*, 116:396–410, 2012.

[45] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[46] Ding-Yun Chen, Xiao-Pei Tian, Yu-Te Shen, and Ming Ouhyoung. On visual similarity based 3d model retrieval. In *Computer graphics forum*, pages 223–232. Wiley Online Library, 2003.

[47] Hui Chen and Bir Bhanu. 3d free-form object recognition in range images using local surface patches. *Pattern Recognition Letters*, pages 1252–1262, 2007.

[48] Minmin Chen, Jeffrey Pennington, and Samuel Schoenholz. Dynamical isometry and a mean field theory of RNNs: Gating enables signal propagation in recurrent neural networks. In *Proceedings of the International conference on machine learning (ICML)*, volume 80, pages 873–882. PMLR, Jul 2018.

[49] Gong Cheng, Peicheng Zhou, and Junwei Han. Rifd-cnn: Rotation-invariant and fisher discriminative convolutional neural networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2884–2893. IEEE, 2016.

[50] Warren Cheung and Ghassan Hamarneh. N-sift: N-dimensional scale invariant feature transform for matching medical images. In *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 720–723. IEEE, 2007.

[51] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[52] Sungjoon Choi, Qian-Yi Zhou, Stephen Miller, and Vladlen Koltun. A large dataset of object scans. *arXiv:1602.02481*, 2016.

[53] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[54] Chris A Cocosco, Vasken Kollokian, Remi K-S Kwan, G Bruce Pike, and Alan C Evans. Brainweb: Online interface to a 3d mri simulated brain database. In *NeuroImage*. Citeseer, 1997.

[55] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the International conference on machine learning (ICML)*, pages 2990–2999, 2016.

[56] Tim Cooijmans, Nicolas Ballas, César Laurent, Çağlar Gülçehre, and Aaron Courville. Recurrent batch normalization. *arXiv preprint arXiv:1603.09025*, 2016.

[57] Camille Couprie. Multi-label energy minimization for object class segmentation. In *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 2233–2237. IEEE, 2012.

[58] Camille Couprie, Clément Farabet, Laurent Najman, and Yann LeCun. Indoor semantic segmentation using depth information. *arXiv preprint arXiv:1301.3572*, 2013.

[59] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In *Proceedings of*

the IEEE conference on computer vision and pattern recognition (CVPR), pages 113–123, 2019.

[60] John D. and Anderson Jr. *Computational Fluid Dynamics*. McGraw-Hill, 1995.

[61] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 5828–5839. IEEE, 2017.

[62] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 886–893. IEEE, 2005.

[63] Tal Darom and Yosi Keller. Scale-invariant features for 3-d mesh models. *IEEE Transactions on Image Processing*, 21:2758–2769, 2012.

[64] Liuyuan Deng, Ming Yang, Tianyi Li, Yuesheng He, and Chunxiang Wang. Rfbnet: Deep multimodal networks with residual fusion blocks for rgb-d semantic segmentation. *arXiv preprint arXiv:1907.00135*, 2019.

[65] Zhuo Deng, Sinisa Todorovic, and Longin Jan Latecki. Semantic segmentation of rgbd images with mutex constraints. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1733–1741. IEEE, 2015.

[66] H Quynh Dinh and Liefei Xu. Measuring the similarity of vector fields using global distributions. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 187–196. Springer, 2008.

[67] Erickson R do Nascimento, Gabriel L Oliveira, Antônio W Vieira, and Mario FM Campos. On the development of a robust, fast and lightweight keypoint descriptor. *Neurocomputing*, 120:141–155, 2013.

[68] Piotr Dollár, Vincent Rabaud, Garrison Cottrell, and Serge Belongie. Behavior recognition via sparse spatio-temporal features. In *Workshop on visual surveillance and performance evaluation of tracking and surveillance.*, pages 65–72. IEEE, 2005.

[69] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3d fully convolutional networks for subcortical segmentation in mri: A large-scale study. *NeuroImage*, 170:456–470, 2017.

[70] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2625–2634. IEEE, 2015.

[71] Andreas Doumanoglou, Rigas Kouskouridas, Sotiris Malassiotis, and Tae-Kyun Kim. Recovering 6d object pose and predicting next-best-view in the crowd. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3583–3592. IEEE, 2016.

[72] Bertram Drost, Markus Ulrich, Nassir Navab, and Slobodan Ilic. Model globally, match locally: Efficient and robust 3d object recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE, 2010.

[73] Yong Du, Wei Wang, and Liang Wang. Hierarchical recurrent neural network for skeleton based action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1110–1118. IEEE, 2015.

[74] Aman Dureja and Payal Pahwa. Image retrieval techniques: a survey. *International Journal of Engineering & Technology*, 7(1.2):215–219, 2018.

[75] Julia Ebling and Gerik Scheuermann. Clifford convolution and pattern matching on vector fields. In *Proceedings of the 14th IEEE visualization*, page 26. IEEE Computer Society, 2003.

[76] Alexei A Efros, Alexander C Berg, Greg Mori, and Jitendra Malik. Recognizing action at a distance. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, page 726. IEEE, 2003.

[77] David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 2650–2658. IEEE, 2015.

[78] Andreas Eitel, Jost Tobias Springenberg, Luciano Spinello, Martin Riedmiller, and Wolfram Burgard. Multimodal deep learning for robust rgb-d object recognition. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 681–687. IEEE, 2015.

[79] Hanan ElNaghy, Safwat Hamad, and M Essam Khalifa. Taxonomy for 3d content-based object retrieval methods. *International Journal of Recent Research and Applied Studies (IJRRAS)*, 14:412–446, 2013.

[80] Felix Endres, Jürgen Hess, Nikolas Engelhard, Jürgen Sturm, Daniel Cremers, and Wolfram Burgard. An evaluation of the rgb-d slam system. In *IEEE international conference on robotics and automation (ICRA)*, pages 1691–1696. IEEE, 2012.

[81] Felix Endres, Jürgen Hess, Jürgen Sturm, Daniel Cremers, and Wolfram Burgard. 3-d mapping with an rgb-d camera. *Transactions on Robotics*, 30:177–187, 2014.

[82] Martin Engelcke, Dushyant Rao, Dominic Zeng Wang, Chi Hay Tong, and Ingmar Posner. Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks. In *IEEE international conference on robotics and automation (ICRA)*, pages 1355–1361. IEEE, 2017.

[83] POST F., V. ROLIJK B., H AUSER H., L ARAMEE R., and D OLEISCH H. The state of the art in flow visualisation: Feature extraction and tracking. *Computer Graphics Forum*, 22(4):775–792, 2003.

[84] Yuchen Fan, Yao Qian, Feng-Long Xie, and Frank K Soong. Tts synthesis with bidirectional lstm based recurrent neural networks. In *Fifteenth annual conference of the international speech communication association*, 2014.

[85] Clément Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. Scene parsing with multiscale feature learning, purity trees, and optimal covers. In *Proceedings of the International conference on machine learning (ICML)*, pages 1857–1864. Omnipress, 2012.

[86] Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 35:1915–1929, 2013.

[87] Christoph Feichtenhofer, Axel Pinz, and Andrew Zisserman. Convolutional two-stream network fusion for video action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1933–1941. IEEE, 2016.

[88] Basura Fernando, Stratis Gavves, Oramas Mogrovejo, José Antonio, Amir Ghodrati, and Tinne Tuytelaars. Modeling video evolution for action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 5378–5387. IEEE, 2015.

[89] Michael Firman. Rgbd datasets: Past, present and future. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 19–31. IEEE, 2016.

[90] Alex Flint, Anthony Dick, and Anton Van Den Hengel. Thrift: Local 3d structure recognition. In *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA)*, pages 182–188. IEEE, 2007.

[91] Andrea Frome, Daniel Huber, Ravi Kolluri, Thomas Bülow, and Jitendra Malik. Recognizing objects in range data using regional point descriptors. In *Proceedings of the European conference on computer vision (ECCV)*, pages 224–237. Springer, 2004.

[92] Jiyang Gao, Chen Sun, Zhenheng Yang, and Ram Nevatia. Tall: Temporal activity localization via language query. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 5267–5275. IEEE, 2017.

[93] Yue Gao, Qionghai Dai, and Nai-Yao Zhang. 3d model comparison using spatial structure circular descriptor. *Pattern Recognition*, 43:1142–1151, 2010.

[94] Noa Garcia. Temporal aggregation of visual features for large-scale image-to-video retrieval. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, pages 489–492. ACM, 2018.

[95] Noa Garcia and George Vogiatzis. Dress like a star: Retrieving fashion products from videos. In *Proceedings of the IEEE international conference on computer vision workshops (ICCVW)*, pages 2293–2299. IEEE, 2017.

[96] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[97] Christoph Garth, Robert S Laramee, Xavier Tricoche, Jürgen Schneider, and Hans Hagen. Extraction and visualization of swirl and tumble motion from engine simulation data. In *Topology-based Methods in Visualization*, pages 121–135. Springer, 2007.

[98] Andreas Geiger. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3354–3361. IEEE, 2012.

[99] Felix A Gers, Nicol N Schraudolph, and Jürgen Schmidhuber. Learning precise timing with lstm recurrent networks. *Journal of machine learning research (JMLR)*, 3:115–143, 2002.

[100] R. B Girshick, P. F Felzenszwalb, and D. A Mcallester. Object detection with grammar models. In *Advances in Neural Information Processing Systems*, pages 442–450, 2011.

[101] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. PMLR, 2011.

[102] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[103] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. In *Proceedings of the International conference on machine learning (ICML)*, pages III–1319–III–1327. Omnipress, 2013.

[104] Raghav Goyal, Samira Ebrahimi Kahou, Vincent Michalski, Joanna Materzynska, Susanne Westphal, Heuna Kim, Valentin Haenel, Ingo Fruend, Peter Yianilos, Moritz Mueller-Freitag, et al. The" something something" video database for learning and evaluating visual common sense. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, page 3. IEEE, 2017.

[105] Lars Graening and Thomas Ramsay. Flow field data mining based on a compact streamline representation. Technical report, SAE Technical Paper, 2015.

[106] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *Transactions on neural networks and learning systems*, 28:2222–2232, 2017.

[107] Wulong Guo, Weiduo Hu, Chang Liu, and Tingting Lu. 3d object recognition from cluttered and occluded scenes with a compact local feature. *Machine vision and applications*, 30:763–783, 2019.

[108] Xiaoxiao Guo, Wei Li, and Francesco Iorio. Convolutional neural networks for steady flow approximation. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 481–490. ACM, 2016.

[109] Y Guo, Ferdous Sohel, Mohammed Bennamoun, M Lu, and J Wan. Trisi: A distinctive local surface descriptor for 3d modeling and object recognition. In *8th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 86–93. Scitepress, 2013.

[110] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.

[111] Yanming Guo, Yu Liu, **Theodoros Georgiou**, and Michael Lew. A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval (IJMIR)*, 7:87–93, 2018.

[112] Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, and Jianwei Wan. 3d object recognition in cluttered scenes with local surface features: a survey. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, pages 2270–2287, 2014.

[113] Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Min Lu, and Jianwei Wan. Rotational projection statistics for 3d local surface description and object recognition. *International journal of computer vision (IJCV)*, 105:63–86, 2013.

[114] Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Jianwei Wan, and Min Lu. A novel local surface feature for 3d object recognition under clutter and occlusion. *Information Sciences*, pages 196–213, 2015.

[115] Saurabh Gupta, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. Indoor scene understanding with rgb-d images: Bottom-up segmentation, object detection and semantic segmentation. *International journal of computer vision (IJCV)*, 112:133–149, 2015.

[116] Saurabh Gupta, Pablo Arbelaez, and Jitendra Malik. Perceptual organization and recognition of indoor scenes from rgb-d images. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 564–571. IEEE, 2013.

[117] Saurabh Gupta, Ross Girshick, Pablo Arbeláez, and Jitendra Malik. Learning rich features from rgb-d images for object detection and segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 345–360. Springer, 2014.

[118] Simon Hadfield, Karel Lebeda, and Richard Bowden. Hollywood 3d: What are the best 3d features for action recognition? *International journal of computer vision (IJCV)*, 121:95–110, 2017.

[119] Dongyoon Han, Jiwhan Kim, and Junmo Kim. Deep pyramidal residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 5927–5935, 2017.

[120] Ju Han and Kai-Kuang Ma. Fuzzy color histogram and its use in color image retrieval. *IEEE Transactions on image Processing*, 11(8):944–952, 2002.

[121] Ankur Handa, Viorica Patraucean, Vijay Badrinarayanan, Simon Stent, and Roberto Cipolla. Understanding real world indoor scenes with synthetic data. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 4077–4085. IEEE, 2016.

[122] Mehrtash Harandi and Basura Fernando. Generalized backpropagation,\'{E} tude de cas: Orthogonality. *arXiv preprint arXiv:1611.05927*, 2016.

[123] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, pages 10–5244. Citeseer, 1988.

[124] Tal Hassner. A critical review of action recognition benchmarks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 245–250. IEEE, 2013.

[125] Caner Hazirbas, Lingni Ma, Csaba Domokos, and Daniel Cremers. Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture. In *Asian conference on computer vision (ACCV)*, pages 213–228. Springer, 2016.

[126] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1026–1034. IEEE, 2015.

[127] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 770–778. IEEE, 2016.

[128] Vishakh Hegde and Reza Zadeh. Fusionnet: 3d object classification using multiple data representations. *arXiv preprint arXiv:1607.05695*, 2016.

[129] Geremy Heitz and Daphne Koller. Learning spatial context: Using stuff to find things. In *Proceedings of the European conference on computer vision (ECCV)*, pages 30–43. Springer, 2008.

[130] James Helman and Lambertus Hesselink. Representation and display of vector field topology in fluid flow data sets. *IEEE computer*, 22(8):27–36, 1989.

[131] Joao F Henriques and Andrea Vedaldi. Warped convolutions: Efficient invariance to spatial transformations. In *Proceedings of the International conference on machine learning (ICML)*, pages 1461–1469. PMLR, 2017.

[132] Samitha Herath, Mehrtash Harandi, and Fatih Porikli. Going deeper into action recognition: A survey. *Image and vision computing*, 60:4–21, 2017.

[133] Alexander Hermans, Georgios Floros, and Bastian Leibe. Dense 3d semantic mapping of indoor scenes from rgb-d images. In *IEEE international conference on robotics and automation (ICRA)*, pages 2631–2638. IEEE, 2014.

[134] Stefan Hinterstoisser, Cedric Cagniart, Slobodan Ilic, Peter Sturm, Nassir Navab, Pascal Fua, and Vincent Lepetit. Gradient response maps for real-time detection of textureless objects. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, pages 876–888, 2012.

[135] Stefan Hinterstoisser, Stefan Holzer, Cedric Cagniart, Slobodan Ilic, Kurt Konolige, Nassir Navab, and Vincent Lepetit. Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 858–865. IEEE, 2011.

[136] Stefan Hinterstoisser, Vincent Lepetit, Slobodan Ilic, Stefan Holzer, Gary Bradski, Kurt Konolige, and Nassir Navab. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. In *Asian conference on computer vision (ACCV)*, pages 548–562. Springer, 2012.

[137] Stefan Hinterstoisser, Vincent Lepetit, Naresh Rajkumar, and Kurt Konolige. Going further with point pair features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 834–848. Springer, 2016.

[138] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18:1527–1554, 2006.

[139] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, pages 504–507, 2006.

[140] Geoffrey E Hinton and Terrence J Sejnowski. Learning and releaming in boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1:2, 1986.

[141] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, pages 1735–1780, 1997.

[142] Nico Höft, Hannes Schulz, and Sven Behnke. Fast semantic segmentation of rgb-d scenes with gpu-accelerated deep neural networks. In *Joint German/Austrian Conference on Artificial Intelligence*, pages 80–85. Springer, 2014.

[143] David R Holmes, Ellis L Workman, and Richard A Robb. The nlm-mayo image collection: Common access to uncommon data. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) Workshop*, 2005.

[144] Berthold Klaus Paul Horn. Extended gaussian images. *Proceedings of the IEEE*, pages 1671–1686, 1984.

[145] Binh-Son Hua, Quang-Hieu Pham, Duc Thanh Nguyen, Minh-Khoi Tran, Lap-Fai Yu, and Sai-Kit Yeung. Scenenn: A scene meshes dataset with annotations. In *Fourth International Conference on 3D Vision (3DV)*, pages 92–101. IEEE, 2016.

[146] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2261–2269. IEEE, 2017.

[147] Lei Huang, Li Liu, Fan Zhu, Diwen Wan, Zehuan Yuan, Bo Li, and Ling Shao. Controllable orthogonalization in training dnns. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6429–6438, 2020.

[148] Lei Huang, Xianglong Liu, Bo Lang, and Bo Li. Projection based weight normalization for deep neural networks. *arXiv preprint arXiv:1710.02338*, 2017.

[149] Lei Huang, Xianglong Liu, Bo Lang, Adams Wei Yu, Yongliang Wang, and Bo Li. Orthogonal weight normalization: Solution to optimization over multiple dependent stiefel manifolds in deep neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[150] Lei Huang, Dawei Yang, Bo Lang, and Jia Deng. Decorrelated batch normalization. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 791–800. IEEE, 2018.

[151] Haroon Idrees, Amir R Zamir, Yu-Gang Jiang, Alex Gorban, Ivan Laptev, Rahul Sukthankar, and Mubarak Shah. The thumos challenge on action recognition for videos in the wild. *Computer vision and image understanding (CVIU)*, 155:1–23, 2017.

[152] Anastasia Ioannidou, Elisavet Chatzilari, Spiros Nikolopoulos, and Ioannis Kompatsiaris. Deep learning advances in computer vision with 3d data: A survey. *ACM Computing Surveys (CSUR)*, page 20, 2017.

[153] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, volume 37, pages 448–456. PMLR, 2015.

[154] Sergey Ioffe and Christian Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 448–456. Omnipress, 2015.

[155] L. Helman J. and Hesselink L. Representation and display of vector field topology in fluid flow data sets. *IEEE Computer*, 22(8):27–36, 1989.

[156] L. Helman J. and Hesselink L. Visualizing vector field topology in fluid flows. *IEEE Computer Graphics and Applications*, 11(3):36–46, 1991.

[157] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.

[158] Allison Janoch, Sergey Karayev, Yangqing Jia, Jonathan T Barron, Mario Fritz, Kate Saenko, and Trevor Darrell. A category-level 3d object dataset: Putting the kinect to work. In Andrea Fossati, Juergen Gall, Helmut Grabner, Xiaofeng Ren, and Kurt Konolige, editors, *Consumer Depth Cameras for Computer Vision*, pages 141–165. Springer, 2013.

[159] Kevin Jarrett, Koray Kavukcuoglu, Yann LeCun, et al. What is the best multi-stage architecture for object recognition? In *IEEE 12th international conference on computer vision (ICCV)*, pages 2146–2153. IEEE, 2009.

[160] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, pages 221–231, 2013.

[161] Yu-Gang Jiang, Zuxuan Wu, Jun Wang, Xiangyang Xue, and Shih-Fu Chang. Exploiting feature and class relationships in video categorization with regularized deep neural networks. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 40:352–364, 2018.

[162] Yun Jiang, Stephen Moseson, and Ashutosh Saxena. Efficient grasping from rgbd images: Learning using a new rectangle representation. In *IEEE international conference on robotics and automation (ICRA)*, pages 3304–3311. IEEE, 2011.

[163] Xiaojie Jin, Chunyan Xu, Jiashi Feng, Yunchao Wei, Junjun Xiong, and Shuicheng Yan. Deep learning with s-shaped rectified linear activation units. In *AAAI Conference on Artificial Intelligence*, pages 1737–1743, 2016.

[164] Andrew E. Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 21:433–449, 1999.

[165] Andrew E Johnson, Allan R Klumpp, James B Collier, and Aron A Wolf. Lidar-based hazard avoidance for safe landing on mars. *Journal of guidance, control, and dynamics*, pages 1091–1099, 2002.

[166] Andrew Edie Johnson and Martial Hebert. Surface matching for object recognition in complex three-dimensional scenes. *Image and vision computing*, 16:635–651, 1998.

[167] T. Kadir and M. Brady. Scale saliency: a novel approach to salient feature and scale selection. In *International conference on visual information engineering (VIE)*, pages 25–28. IET, 2003.

[168] Soo Min Kang and Richard P Wildes. Review of action recognition and detection methods. *arXiv preprint arXiv:1610.06906*, 2016.

[169] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1725–1732. IEEE, 2014.

[170] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, et al. The kinetics human action video dataset. *arXiv preprint arXiv:1705.06950*, 2017.

[171] Yan Ke, Rahul Sukthankar, and Martial Hebert. Efficient visual event detection using volumetric features. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 166–173. IEEE, 2005.

[172] Wadim Kehl, Fausto Milletari, Federico Tombari, Slobodan Ilic, and Nassir Navab. Deep learning of local rgb-d patches for 3d object detection and 6d pose estimation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 205–220. Springer, 2016.

[173] Christian Kerl, Jurgen Sturm, and Daniel Cremers. Dense visual slam for rgb-d cameras. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 2100–2106. IEEE, 2013.

[174] Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Geometry driven semantic labeling of indoor scenes. In *Proceedings of the European conference on computer vision (ECCV)*, pages 679–694. Springer, 2014.

[175] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[176] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In *Advances in neural information processing systems 30*, pages 971–980. Curran Associates, Inc., 2017.

[177] Alexander Klaser, Marcin Marszałek, and Cordelia Schmid. A spatio-temporal descriptor based on 3d-gradients. In *Proceedings of the British machine vision conference (BMVC)*, pages 995–1004. BMVA Press, 2008.

[178] Jan Knopp, Mukta Prasad, Geert Willems, Radu Timofte, and Luc Van Gool. Hough transform and 3d surf for robust three dimensional classification. In *Proceedings of the European conference on computer vision (ECCV)*, pages 589–602. Springer, 2010.

[179] Jan J Koenderink and Andrea J van Doorn. Representation of local geometry in the visual system. *Biological cybernetics*, 55:367–375, 1987.

[180] Hema S Koppula, Abhishek Anand, Thorsten Joachims, and Ashutosh Saxena. Semantic labeling of 3d point clouds for indoor scenes. In *Advances in neural information processing systems 24*, pages 244–252. Curran Associates, Inc., 2011.

[181] Iryna Korshunova, Wenzhe Shi, Joni Dambre, and Lucas Theis. Fast face-swap using convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[182] Adriana Kovashka and Kristen Grauman. Learning a hierarchy of discriminative space-time neighborhood features for human action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2046–2053. IEEE, 2010.

[183] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[184] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[185] Alexander Krull, Eric Brachmann, Frank Michel, Michael Ying Yang, Stefan Gumhold, and Carsten Rother. Learning analysis-by-synthesis for 6d pose estimation in rgb-d images. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 954–962. IEEE, 2015.

[186] Hildegard Kuehne, Hueihan Jhuang, Estíbaliz Garrote, Tomaso Poggio, and Thomas Serre. Hmdb: a large video database for human motion recognition. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 2556–2563. IEEE, 2011.

[187] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *IEEE international conference on robotics and automation (ICRA)*, pages 1817–1824. IEEE, 2011.

[188] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. Rgb-d object recognition: Features, algorithms, and a large scale benchmark. In *Consumer Depth Cameras for Computer Vision*, pages 167–192. Springer, 2013.

[189] Dmitry Laptev, Nikolay Savinov, Joachim M Buhmann, and Marc Pollefeys. Ti-pooling: transformation-invariant pooling for feature learning in convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 289–297, 2016.

[190] Ivan Laptev. On space-time interest points. *International journal of computer vision (IJCV)*, pages 107–123, 2005.

[191] Ivan Laptev, Barbara Caputo, Christian Schüldt, and Tony Lindeberg. Local velocity-adapted motion events for spatio-temporal recognition. *Computer vision and image understanding (CVIU)*, 108:207–229, 2007.

[192] Ivan Laptev and Tony Lindeberg. Velocity adaptation of space-time interest points. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 52–56. IEEE, 2004.

[193] Ivan Laptev and Tony Lindeberg. Local descriptors for spatio-temporal recognition. In *Spatial Coherence for Visual Motion Analysis*, pages 91–103. Springer, 2006.

[194] Ivan Laptev, Marcin Marszalek, Cordelia Schmid, and Benjamin Rozenfeld. Learning realistic human actions from movies. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1–8. IEEE, 2008.

[195] Graciela Lara López, Adriana Pena Pérez Negrón, Angélica De Antonio Jiménez, Jaime Ramírez Rodríguez, and Ricardo Imbert Paredes. Comparative analysis of shape descriptors for 3d objects. *Multimedia Tools and Applications*, 76:6993–7040, 2017.

[196] Robert S Laramee, Helwig Hauser, Lingxiao Zhao, and Frits H Post. Topology-based flow visualization, the state of the art. In *Topology-based methods in visualization*, pages 1–19. Springer, 2007.

[197] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the International conference on machine learning (ICML)*, pages 473–480. ACM, 2007.

[198] César Laurent, Gabriel Pereyra, Philémon Brakel, Ying Zhang, and Yoshua Bengio. Batch normalized recurrent neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2657–2661. IEEE, 2016.

[199] Yann LeCun, Bernhard Boser, John Denker, Donnie Henderson, R Howard, Wayne Hubbard, and Lawrence Jackel. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2:396–404, 1989.

[200] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[201] Yann LeCun, Lawrence D Jackel, Léon Bottou, Corinna Cortes, John S Denker, Harris Drucker, Isabelle Guyon, Urs A Muller, Eduard Sackinger, Patrice Simard, et al. Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, 261:276, 1995.

[202] Yann LeCun, D Touresky, G Hinton, and T Sejnowski. A theoretical framework for back-propagation. In *Proceedings of the 1988 connectionist models summer school*, volume 1, pages 21–28. CMU, Pittsburgh, Pa: Morgan Kaufmann, 1988.

[203] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570. PMLR, 2015.

[204] Ian Lenz, Honglak Lee, and Ashutosh Saxena. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research*, 34(4-5):705–724, 2015.

[205] S. Leutenegger, M. Chli, and R. Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 2548–2555, 2011.

[206] Bo Li, Yijuan Lu, Chunyuan Li, Afzal Godil, Tobias Schreck, Masaki Aono, Martin Burtscher, Hongbo Fu, Takahiko Furuya, Henry Johan, et al. Shrec14 track: extended large scale sketch-based 3d shape retrieval. In *Eurographics workshop on 3D object retrieval (3DOR)*, pages 121–130, 2014.

[207] Bo Li, Tianlei Zhang, and Tian Xia. Vehicle detection from 3d lidar using fully convolutional network. *arXiv preprint arXiv:1608.07916*, 2016.

[208] Wanqing Li, Zhengyou Zhang, and Zicheng Liu. Action recognition based on a bag of 3d points. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 9–14. IEEE, 2010.

[209] Y. Li, S. Wang, Q. Tian, and X. Ding. A survey of recent advances in visual feature detection. *Neurocomputing*, 149:736–751, 2015.

[210] Yanshan Li, Rongjie Xia, Qinghua Huang, Weixin Xie, and Xuelong Li. Survey of spatio-temporal interest point detection algorithms in video. *IEEE Access*, 5:10323–10331, 2017.

[211] Yanshan Li, Rongjie Xia, and Weixin Xie. A unified model of appearance and motion of video and its application in stip detection. *Signal, Image and Video Processing*, pages 403–410, 2018.

[212] Zhen Li, Yukang Gan, Xiaodan Liang, Yizhou Yu, Hui Cheng, and Liang Lin. Lstm-cf: Unifying context modeling and fusion with lstms for rgb-d scene labeling. In *Proceedings of the European conference on computer vision (ECCV)*, pages 541–557. Springer, 2016.

[213] Zhen Li, Yukang Gan, Xiaodan Liang, Yizhou Yu, Hui Cheng, and Liang Lin. Rgb-d scene labeling with long short-term memorized fusion model. *arXiv preprint arXiv:1604.05000*, 2016.

[214] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast autoaugment. In *Advances in Neural Information Processing Systems*, pages 6665–6675, 2019.

[215] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE, 2017.

[216] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[217] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Proceedings of the European conference on computer vision (ECCV)*, pages 740–755. Springer, 2014.

[218] Julia Ling, Andrew Kurzawski, and Jeremy Templeton. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, 807:155–166, 2016.

[219] Jun Liu, Amir Shahroudy, Dong Xu, and Gang Wang. Spatio-temporal lstm with trust gates for 3d human action recognition. In *Proceedings of the European conference on computer vision (ECCV)*, pages 816–833. Springer, 2016.

[220] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen. Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition*, 62:135–160, 2017.

[221] Long Liu, Zhixuan Xi, RuiRui Ji, and Weigang Ma. Advanced deep learning techniques for image style transfer: A survey. *Signal Processing: Image Communication*, 78:465–470, 2019.

[222] Yu Liu, Yanming Guo, **Theodoros Georgiou**, and Michael Lew. Fusion that matters: convolutional fusion networks for visual recognition. *Multimedia Tools and Applications*, 77:1–28, 2018.

[223] Tsz-Wai Rachel Lo and J Paul Siebert. Local feature extraction and matching on range images: 2.5 d sift. *Computer vision and image understanding (CVIU)*, pages 1235–1250, 2009.

[224] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3431–3440. IEEE, 2015.

[225] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1150–1157. IEEE, 1999.

[226] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision (IJCV)*, 60:91–110, 2004.

[227] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 1981 DARPA imaging understanding Workshop (IJCAI)*. Vancouver, BC, Canada, 1981.

[228] Minh-Thang Luong, Ilya Sutskever, Quoc V Le, Oriol Vinyals, and Wojciech Zaremba. Addressing the rare word problem in neural machine translation. *arXiv preprint arXiv:1410.8206*, 2014.

[229] Jiang M, Machiraju R, and Thompson D. Detection and visualization of vortices. *The visualization handbook*, page 295, 2005.

[230] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the International Conference on Machine Learning (ICML)*, page 3. Omnipress, 2013.

[231] Chris Maes, Thomas Fabry, Johannes Keustermans, Dirk Smeets, Paul Suetens, and Dirk Vandermeulen. Feature detection on 3d face surfaces for pose normalisation and recognition. In *Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6. IEEE, 2010.

[232] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In *Proceedings of the European conference on computer vision (ECCV)*, pages 183–196, 2010.

[233] D. Marcos, M. Volpi, N. Komodakis, and D. Tuia. Rotation equivariant vector field networks. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 5048–5057, 2017.

[234] Diego Marcos, Michele Volpi, and Devis Tuia. Learning rotation invariant convolutional filters for texture classification. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 2012–2017. IEEE, 2016.

[235] Marcin Marszalek, Ivan Laptev, and Cordelia Schmid. Actions in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2929–2936. IEEE, 2009.

[236] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International conference on artificial neural networks*, pages 52–59. Springer, 2011.

[237] Pyry Matikainen, Martial Hebert, and Rahul Sukthankar. Trajectons: Action recognition through the motion analysis of tracked features. In *Proceedings of the IEEE international conference on computer vision workshops (ICCVW)*, pages 514–521. IEEE, 2009.

[238] Takahiro Matsuda, Takahiko Furuya, and Ryutarou Ohbuchi. Lightweight binary voxel shape features for 3d data matching and retrieval. In *IEEE International Conference on Multimedia Big Data*, pages 100–107. IEEE, 2015.

[239] Daniel Maturana and Sebastian Scherer. 3d convolutional neural networks for landing zone detection from lidar. In *IEEE international conference on robotics and automation (ICRA)*, pages 3471–3478. IEEE, 2015.

[240] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 922–928. IEEE, 2015.

[241] John McCormac, Ankur Handa, Stefan Leutenegger, and Andrew J Davison. Scenenet rgb-d: 5m photorealistic images of synthetic indoor trajectories with ground truth. *arXiv preprint arXiv:1612.05079*, 2016.

[242] Roland Memisevic and Geoffrey Hinton. Unsupervised learning of image transformations. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1–8. IEEE, 2007.

[243] Stefan Menzel and Bernhard Sendhoff. Representing the change - free form deformation for evolutionary design optimisation. In Tina Yu, David Davis, Cem Baydar, and Rajkumar Roy, editors, *Evolutionary Computation in Practice*, chapter 4, pages 63–86. Springer, 2008.

[244] Ross Messing, Chris Pal, and Henry Kautz. Activity recognition using the velocity histories of tracked keypoints. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 104–111. IEEE, 2009.

[245] Ajmal Mian, Mohammed Bennamoun, and Robyn Owens. On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International journal of computer vision (IJCV)*, pages 348–361, 2010.

[246] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *Proceedings of the IEEE international conference on computer vision (ICCV)*, 60(1):63–86, 2004.

[247] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 27:1615–1630, 2005.

[248] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[249] Farzin Mokhtarian, Nasser Khalili, and Peter Yuen. Multi-scale free-form 3d object recognition using 3d models. *Image and Vision Computing*, 19:271–281, 2001.

[250] Mathew Monfort, Alex Andonian, Bolei Zhou, Kandan Ramakrishnan, Sarah Adel Bargal, Tom Yan, Lisa Brown, Quanfu Fan, Dan Gutfreund, Carl Vondrick, et al. Moments in time dataset: one million videos for event understanding. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 42(2):502–508, 2019.

[251] Andreas C Müller and Sven Behnke. Learning depth-sensitive conditional random fields for semantic segmentation of rgb-d images. In *IEEE international conference on robotics and automation (ICRA)*, pages 6232–6237. IEEE, 2014.

[252] Raul Mur-Artal and Juan D Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *Transactions on Robotics*, 33:1255–1262, 2017.

[253] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the International conference on machine learning (ICML)*, pages 807–814. Omnipress, 2010.

[254] Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, and George Toderici. Beyond short snippets: Deep networks for video classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 4694–4702. IEEE, 2015.

[255] Jiquan Ngiam, Zhenghao Chen, Pang W Koh, and Andrew Y Ng. Learning deep energy models. In *Proceedings of the International conference on machine learning (ICML)*, pages 1105–1112. Omnipress, 2011.

[256] Dong Ni, Yim Pan Chui, Yingge Qu, Xuan Yang, Jing Qin, Tien-Tsin Wong, Simon SH Ho, and Pheng Ann Heng. Reconstruction of volumetric ultrasound panorama based on improved 3d sift. *Computerized Medical Imaging and Graphics*, 33:559–566, 2009.

[257] Juan Carlos Niebles, Hongcheng Wang, and Li Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *International journal of computer vision (IJCV)*, 79:299–318, 2008.

[258] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1520–1528. IEEE, 2015.

[259] John Novatnack and Ko Nishino. Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images. In *Proceedings of the European conference on computer vision (ECCV)*, pages 440–453. Springer, 2008.

[260] Eric Nowak, Frédéric Jurie, and Bill Triggs. Sampling strategies for bag-of-features image classification. In *Proceedings of the European conference on computer vision (ECCV)*, pages 490–503. Springer, 2006.

[261] Antonios Oikonomopoulos, Ioannis Patras, and Maja Pantic. Spatiotemporal salient points for visual recognition of human actions. *Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36:710–719, 2005.

[262] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 24:971–987, 2002.

[263] Nuria M Oliver, Barbara Rosario, and Alex P Pentland. A bayesian computer vision system for modeling human interactions. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 22:831–843, 2000.

[264] openFOAM and the openFOAM foundation. openfoam, 2011-2017.

[265] Omar Oreifej and Zicheng Liu. Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 716–723. IEEE, 2013.

[266] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. Shape distributions. *ACM Transactions on Graphics (TOG)*, 21:807–832, 2002.

[267] Mete Ozay and Takayuki Okatani. Optimization on submanifolds of convolution kernels in cnns. *arXiv preprint arXiv:1610.07008*, 2016.

[268] Umut Özaydın, **Theodoros Georgiou**, and Michael Lew. A comparison of cnn and classic features for image retrieval. In *International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–4, 2019.

[269] Chavdar Papazov and Darius Burschka. An efficient ransac for 3d object recognition in noisy and occluded scenes. In *Asian conference on computer vision (ACCV)*, pages 135–148. Springer, 2010.

[270] Chavdar Papazov, Sami Haddadin, Sven Parusel, Kai Krieger, and Darius Burschka. Rigid 3d geometry matching for grasping of known objects in cluttered scenes. *The International Journal of Robotics Research (IJRR)*, pages 538–553, 2012.

[271] Seong-Jin Park, Ki-Sang Hong, and Seungyong Lee. Rdfnet: Rgb-d multi-level residual feature fusion for indoor semantic segmentation. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 4990–4999. IEEE, 2017.

[272] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[273] A. Pobitzer, R. Peikert, R. Fuchs, B. Schindler, A. Kuhn, H. Theisel, K. Matkovi, and H. Hauser. The state of the art in topology-based visualization of unsteady flow. *Computer Graphics forum*, 30:1789–1811, 2011.

[274] Ronald Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28:976–990, 2010.

[275] Frits H Post, Benjamin Vrolijk, Helwig Hauser, Robert S Laramee, and Helmut Doleisch. The state of the art in flow visualisation: Feature extraction and tracking. In *Computer Graphics Forum*, volume 22, pages 775–792. Wiley Online Library, 2003.

[276] Christopher Poultney, Sumit Chopra, Yann L Cun, et al. Efficient learning of sparse representations with an energy-based model. In *Advances in neural information processing systems*, pages 1137–1144, 2007.

[277] Nan Pu, **Theodoros Georgiou**, Erwin M Bakker, and Michael Lew. Learning a domain-invariant embedding for unsupervised person re-identification. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019.

[278] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgb-d data. *arXiv preprint arXiv:1711.08488*, 2017.

[279] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. IEEE, 2017.

[280] Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 5648–5656. IEEE, 2016.

[281] Xiaojuan Qi, Renjie Liao, Jiaya Jia, Sanja Fidler, and Raquel Urtasun. 3d graph neural networks for rgbd semantic segmentation. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 5199–5208. IEEE, 2017.

[282] Alastair Quadros, James Patrick Underwood, and Bertrand Douillard. Sydney urban objects dataset. `http://www.acfr.usyd.\protect\discretionary{\char\hyphenchar\font}{}{}edu.au/papers/SydneyUrbanObjectsDataset.shtml.`, 2013.

[283] Siwen Quan, Jie Ma, Tao Ma, Fangyu Hu, and Bin Fang. Representing local shape geometry from multi-view silhouette perspective: A distinctive and robust binary 3d feature. *Signal Processing: Image Communication*, 65:67–80, 2018.

[284] Hossein Rahmani, Arif Mahmood, Du Huynh, and Ajmal Mian. Histogram of oriented principal components for cross-view action recognition. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 38:2430–2443, 2016.

[285] Hossein Rahmani, Arif Mahmood, Du Q Huynh, and Ajmal Mian. Hopc: Histogram of oriented principal components of 3d pointclouds for action recognition. In *Proceedings of the European conference on computer vision (ECCV)*, pages 742–757. Springer, 2014.

[286] Michaela Regneri, Marcus Rohrbach, Dominikus Wetzel, Stefan Thater, Bernt Schiele, and Manfred Pinkal. Grounding action descriptions in videos. *Transactions of the Association for Computational Linguistics*, 1:25–36, 2013.

[287] Mengye Ren, Renjie Liao, Raquel Urtasun, Fabian H Sinz, and Richard S Zemel. Normalizing the normalizers: Comparing and extending network normalization schemes. *arXiv preprint arXiv:1611.04520*, 2016.

[288] Xiaofeng Ren, Liefeng Bo, and Dieter Fox. Rgb-(d) scene labeling: Features and algorithms. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2759–2766. IEEE, 2012.

[289] Colin Rennie, Rahul Shome, Kostas E Bekris, and Alberto F De Souza. A dataset for improved rgbd-based object detection and pose estimation for warehouse pick-and-place. *Robotics and Automation Letters*, 1:1179–1185, 2016.

[290] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *Proceedings of the European conference on computer vision (ECCV)*, pages 102–118. Springer, 2016.

[291] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. Contractive auto-encoders: Explicit invariance during feature extraction. In *Proceedings of the International conference on machine learning (ICML)*, pages 833–840. Omnipress, 2011.

[292] Reyes Rios-Cabrera and Tinne Tuytelaars. Discriminatively trained templates for 3d object detection: A real time scalable approach. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 2048–2055. IEEE, 2013.

[293] S. Laramee Robert, Hauser Helwig, xiao Ling, Frits Zhao, and Post H. Topology-based flow visualization, the state of the art. *Hauser H., Hagen H., Theisel H. (eds) Topology-based Methods in Visualization. Mathematics and Visualization.*, pages 1–19, 2007.

[294] Stephen K Robinson. Coherent motions in the turbulent boundary layer. *Annual Review of Fluid Mechanics*, 23(1):601–639, 1991.

[295] Mikel D Rodriguez, Javed Ahmed, and Mubarak Shah. Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1–8. IEEE, 2008.

[296] Karl Rohr. On 3d differential operators for detecting point landmarks. *Image and Vision Computing*, 15:219–233, 1997.

[297] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3234–3243. IEEE, 2016.

[298] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *Proceedings of the European conference on computer vision (ECCV)*, pages 430–443. Springer, 2006.

[299] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary R Bradski. Orb: An efficient alternative to sift or surf. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, volume 11, page 2. Citeseer, 2011.

[300] Sebastian Ruder, Matthew E Peters, Swabha Swayamdipta, and Thomas Wolf. Transfer learning in natural language processing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, pages 15–18, 2019.

[301] DE Rumelhart. Hinton and williams, rj (1986):learning internal representations by error propagation. *parallel distributed processing*, 1, 1986.

[302] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision (IJCV)*, 115:211–252, 2015.

[303] Raif M Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, pages 225–233. Eurographics Association, 2007.

[304] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3d registration. In *IEEE international conference on robotics and automation (ICRA)*, pages 3212–3217. IEEE, 2009.

[305] Radu Bogdan Rusu, Nico Blodow, Zoltan Csaba Marton, and Michael Beetz. Aligning point cloud views using persistent feature histograms. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 3384–3391. IEEE, 2008.

[306] Ajmal Saeed Mian, Mohammed Bennamoun, and Robyn Owens. Automated 3d model-based free-form object recognition. *Sensor Review*, pages 206–215, 2004.

[307] E. Salahat and M. Qasaimeh. Recent advances in features extraction and description algorithms: A comprehensive survey. In *IEEE international conference on industrial technology (ICIT)*, pages 1059–1063, 2017.

[308] Ruslan Salakhutdinov. Learning and evaluating boltzmann machines. *Tech. Rep., Technical Report UTML TR 2008-002, Department of Computer Science, University of Toronto*, 2008.

[309] Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455. PMLR, 2009.

[310] Ruslan Salakhutdinov and Hugo Larochelle. Efficient learning of deep boltzmann machines. In *Artificial intelligence and statistics*, pages 693–700. PMLR, 2010.

[311] Tim Salimans and Diederik P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems 29*, pages 901–909. Curran Associates, Inc., 2016.

[312] Muhamad Risqi U Saputra, Andrew Markham, and Niki Trigoni. Visual slam and structure from motion in dynamic environments: A survey. *ACM Computing Surveys (CSUR)*, page 37, 2018.

[313] Silvio Savarese and Li Fei-Fei. 3d generic object categorization, localization and pose estimation. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1–8. IEEE, 2007.

[314] Manolis Savva, Angel X. Chang, and Pat Hanrahan. Semantically-enriched 3d models for common-sense knowledge. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 24–31. IEEE, 2015.

[315] Sebastian Scherer, Lyle Chamberlain, and Sanjiv Singh. Autonomous landing at unprepared sites by a full-scale helicopter. *Robotics and Autonomous Systems*, pages 1545–1562, 2012.

[316] Gerik Scheuermann. Topological vector field visualization with clifford algebra. In *Ausgezeichnete Informatikdissertationen*, pages 213–222. Springer, 2000.

[317] J. L. Schonberger, H. Hardmeier, T. Sattler, and M. Pollefeys. Comparative evaluation of hand-crafted and learned local features. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1482–1491, 2017.

[318] Christian Schuldt, Ivan Laptev, and Barbara Caputo. Recognizing human actions: a local svm approach. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 32–36. IEEE, 2004.

[319] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *Transactions on Signal Processing*, 45:2673–2681, 1997.

[320] Brent Schwarz. Lidar: Mapping the world in 3d. *Nature Photonics*, page 429, 2010.

[321] Paul Scovanner, Saad Ali, and Mubarak Shah. A 3-dimensional sift descriptor and its application to action recognition. In *Proceedings of the 15th ACM international conference on Multimedia (ICM)*, pages 357–360. ACM, 2007.

[322] Nicu Sebe, Michael Lew, and Thomas S Huang. The state-of-the-art in human-computer interaction. In *International workshop on computer vision in human-computer interaction*, pages 1–6. Springer, 2004.

[323] Nima Sedaghat, Mohammadreza Zolfaghari, Ehsan Amiri, and Thomas Brox. Orientation-boosted voxel nets for 3d object recognition. *arXiv preprint arXiv:1604.03351*, 2016.

[324] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1010–1019. IEEE, 2016.

[325] Eli Shechtman and Michal Irani. Space-time behavior based correlation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 405–412. IEEE, 2005.

[326] Eli Shechtman and Michal Irani. Space-time behavior-based correlation-or-how to tell if two underlying motion fields are similar without computing them? *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 29:2045–2056, 2007.

[327] Baoguang Shi, Song Bai, Zhichao Zhou, and Xiang Bai. Deeppano: Deep panoramic representation for 3-d shape recognition. *Signal Processing Letters*, 22:2339–2343, 2015.

[328] Jianbo Shi et al. Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pages 593–600. IEEE, 1994.

[329] Jau-Ling Shih, Chang-Hsing Lee, and Jian Tang Wang. A new 3d model retrieval approach based on the elevation descriptor. *Pattern Recognition*, 40:283–295, 2007.

[330] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser. The princeton shape benchmark. In *Shape modeling applications, Proceedings*, pages 167–178. IEEE, 2004.

[331] Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244, 2000.

[332] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1297–1304. IEEE, 2011.

[333] Daniel Sieger, Sergius Gaulik, Jascha Achenbach, Stefan Menzel, and Mario Botsch. Constrained space deformation techniques for design optimization. *Computer-Aided Design*, 72:40–51, March 2016.

[334] Nathan Silberman and Rob Fergus. Indoor scene segmentation using a structured light sensor. In *Proceedings of the IEEE international conference on computer vision workshops (ICCVW)*, pages 601–608. IEEE, 2011.

[335] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Proceedings of the European conference on computer vision (ECCV)*, pages 746–760. Springer, 2012.

[336] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[337] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems 27*, pages 568–576. Curran Associates, Inc., 2014.

[338] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning*

Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[339] Arjun Singh, James Sha, Karthik S Narayan, Tudor Achim, and Pieter Abbeel. Bigbird: A large-scale 3d database of object instances. In *IEEE international conference on robotics and automation (ICRA)*, pages 509–516. IEEE, 2014.

[340] Tej Singh and Dinesh Kumar Vishwakarma. Video benchmarks of human action datasets: a review. *Artificial Intelligence Review*, 52:1107–1154, 2019.

[341] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1470–1478, 2003.

[342] Richard Socher, Brody Huval, Bharath Putta Bath, Christopher D Manning, and Andrew Y Ng. Convolutional-recursive deep learning for 3d object classification. In *Advances in neural information processing systems 25*, page 8. Curran Associates Inc., 2012.

[343] Jure Sokolić, Raja Giryes, Guillermo Sapiro, and Miguel RD Rodrigues. Robust large margin deep neural networks. *IEEE Transactions on Signal Processing*, 65(16):4265–4280, 2017.

[344] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 567–576. IEEE, 2015.

[345] Shuran Song and Jianxiong Xiao. Sliding shapes for 3d object detection in depth images. In *Proceedings of the European conference on computer vision (ECCV)*, pages 634–651. Springer, 2014.

[346] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. Semantic scene completion from a single depth image. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1746–1754. IEEE, 2017.

[347] Yale Song, Louis-Philippe Morency, and Randall Davis. Action recognition by hierarchical sequence summarization. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3562–3569. IEEE, 2013.

[348] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*, 2012.

[349] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015.

[350] Hauke Strasdat, Andrew J Davison, JM Martìnez Montiel, and Kurt Konolige. Double window optimisation for constant time visual slam. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 2352–2359. IEEE, 2011.

[351] Jörg Stückler, Nenad Biresev, and Sven Behnke. Semantic mapping using object-class segmentation of rgb-d images. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 3005–3010. IEEE, 2012.

[352] Jörg Stückler, Benedikt Waldvogel, Hannes Schulz, and Sven Behnke. Dense real-time mapping of object-class semantics from rgb-d video. *Journal of Real-Time Image Processing*, 10:599–609, 2015.

[353] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 945–953. IEEE, 2015.

[354] Deqing Sun, Stefan Roth, and Michael J Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International journal of computer vision (IJCV)*, 106:115–137, 2014.

[355] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer graphics forum*, pages 1383–1392. Wiley Online Library, 2009.

[356] Ju Sun, Xiao Wu, Shuicheng Yan, Loong-Fah Cheong, Tat-Seng Chua, and Jintao Li. Hierarchical spatio-temporal context modeling for action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2004–2011. IEEE, 2009.

[357] Michael J Swain and Dana H Ballard. Color indexing. *International journal of computer vision*, 7(1):11–32, 1991.

[358] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017.

[359] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich,

et al. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1–9. IEEE, 2015.

[360] Salzbrunn T., Janicke H., Wischgoll T., and Scheuermann G. The state of the art in flow visualization: Partition-based techniques. *Proceedings of the 2008 Simulation and Visualization Conference*, pages 75–92, 2008.

[361] Shuai Tang, Xiaoyu Wang, Xutao Lv, Tony X Han, James Keller, Zhihai He, Marjorie Skubic, and Shihong Lao. Histogram of oriented normal vectors for object recognition with a depth sensor. In *Asian conference on computer vision (ACCV)*, pages 525–538. Springer, 2012.

[362] Johan WH Tangelder and Remco C Veltkamp. A survey of content based 3d shape retrieval methods. *Multimedia tools and applications*, 39(3):441–471, 2008.

[363] Graham W Taylor, Rob Fergus, Yann LeCun, and Christoph Bregler. Convolutional learning of spatio-temporal features. In *Proceedings of the European conference on computer vision (ECCV)*, pages 140–153. Springer, 2010.

[364] Alex Teichman, Jesse Levinson, and Sebastian Thrun. Towards 3d object recognition via classification of arbitrary object tracks. In *IEEE international conference on robotics and automation (ICRA)*, pages 4034–4041. IEEE, 2011.

[365] Alex Teichman and Sebastian Thrun. Tracking-based semi-supervised learning. *The International Journal of Robotics Research (IJRR)*, 31:804–818, 2012.

[366] Alykhan Tejani, Rigas Kouskouridas, Andreas Doumanoglou, Danhang Tang, and Tae-Kyun Kim. Latent-class hough forests for 6 dof object pose estimation. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 40:119–132, 2017.

[367] Alykhan Tejani, Rigas Kouskouridas, Andreas Doumanoglou, Danhang Tang, and Tae-Kyun Kim. Latent-class hough forests for 6 dof object pose estimation. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 40:119–132, 2018.

[368] **Theodoros Georgiou**, Yu Liu, Wei Chen, and Michael Lew. A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision. *International Journal of Multimedia Information Retrieval (IJMIR)*, pages 1–36, 2019.

[369] **Theodoros Georgiou**, Sebastian Schmitt, Thomas Bäck, and Michael Lew. Orientational equivariant neural networks using clifford convolutions.

[370] **Theodoros Georgiou**, Sebastian Schmitt, Wei Chen, Thomas Bäck, and Michael Lew. Norm loss: An efficient yet effective regularization method for deep neural networks. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE.

[371] **Theodoros Georgiou**, Sebastian Schmitt, Markus Olhofer, Yu Liu, Thomas Bäck, and Michael Lew. Learning fluid flows. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[372] **Theodoros Georgiou**, Sebastian Schmitt, Nan Pu, Wei Chen, Thomas Bäck, and Michael Lew. Comparison of deep learning and hand crafted features for mining simulation data. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*. IEEE.

[373] Bart Thomee, Mark J Huiskes, Erwin Bakker, and Michael Lew. Large scale image copy detection evaluation. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval (ICMIR)*, pages 59–66. ACM, 2008.

[374] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. The new data and new challenges in multimedia research. *arXiv preprint arXiv:1503.01817*, 2015.

[375] Federico Tombari and Luigi Di Stefano. Hough voting for 3d object recognition under occlusion and clutter. *IPSJ Transactions on Computer Vision and Applications*, pages 20–29, 2012.

[376] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *Proceedings of the European conference on computer vision (ECCV)*, pages 356–369. Springer, 2010.

[377] Federico Tombari, Samuele Salti, and Luigi Di Stefano. A combined texture-shape descriptor for enhanced 3d feature matching. In *18th IEEE international conference on image processing (ICIP)*, pages 809–812. IEEE, 2011.

[378] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Performance evaluation of 3d keypoint detectors. *International journal of computer vision (IJCV)*, 102:198–220, 2013.

[379] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 4489–4497. IEEE, 2015.

[380] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 6450–6459. IEEE, 2018.

[381] Ludovic Trottier, Philippe Gigu, Brahim Chaib-draa, et al. Parametric exponential linear unit for deep convolutional neural networks. In *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 207–214. IEEE, 2017.

[382] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[383] Abhinav Valada, Rohit Mohan, and Wolfram Burgard. Self-supervised model adaptation for multimodal semantic segmentation. *International journal of computer vision (IJCV)*, 2019.

[384] Gul Varol, Ivan Laptev, and Cordelia Schmid. Long-term temporal convolutions for action recognition. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 40:1510–1517, 2017.

[385] Antonio W Vieira, Erickson R Nascimento, Gabriel L Oliveira, Zicheng Liu, and Mario FM Campos. Stop: Space-time occupancy patterns for 3d action recognition from depth map sequences. In *Iberoamerican congress on pattern recognition*, pages 252–259. Springer, 2012.

[386] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the International conference on machine learning (ICML)*, pages 1096–1103. ACM, 2008.

[387] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11:3371–3408, 2010.

[388] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, page 3. IEEE, 2001.

[389] Eugene Vorontsov, Chiheb Trabelsi, Samuel Kadoury, and Chris Pal. On orthogonality and learning recurrent networks with long term dependencies. *arXiv preprint arXiv:1702.00071*, 2017.

[390] Wang W., Wang W., and Li S. From numerics to combinatorics: a survey of topological methods for vector field visualization. *Journal of Visualization*, 19:727–752, 2016.

[391] Anran Wang, Jiwen Lu, Gang Wang, Jianfei Cai, and Tat-Jen Cham. Multi-modal unsupervised feature learning for rgb-d scene labeling. In *Proceedings of the European conference on computer vision (ECCV)*, pages 453–467. Springer, 2014.

[392] Chu Wang, Marcello Pelillo, and Kaleem Siddiqi. Dominant set clustering and pooling for multi-view 3d object recognition. *arXiv preprint arXiv:1906.01592*, 2019.

[393] Dominic Zeng Wang, Ingmar Posner, and Paul Newman. What could move? finding cars, pedestrians and bicyclists in 3d laser data. In *IEEE international conference on robotics and automation (ICRA)*, pages 4038–4044. IEEE, 2012.

[394] Guangrun Wang, Ping Luo, Xinjiang Wang, Liang Lin, et al. Kalman normalization: Normalizing internal representations across network layers. In *Advances in Neural Information Processing Systems 31*, pages 21–31. Curran Associates, Inc., 2018.

[395] Heng Wang, Alexander Kläser, Cordelia Schmid, and Cheng-Lin Liu. Action recognition by dense trajectories. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3169–3176. IEEE, 2011.

[396] Heng Wang, Alexander Kläser, Cordelia Schmid, and Cheng-Lin Liu. Dense trajectories and motion boundary descriptors for action recognition. *International journal of computer vision (IJCV)*, 103:60–79, 2013.

[397] Heng Wang and Cordelia Schmid. Action recognition with improved trajectories. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 3551–3558. IEEE, 2013.

[398] Heng Wang, Muhammad Muneeb Ullah, Alexander Klaser, Ivan Laptev, and Cordelia Schmid. Evaluation of local spatio-temporal features for action recognition. In *Proceedings of the British machine vision conference (BMVC)*, pages 1–11, 2009.

[399] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Learning actionlet ensemble for 3d human action recognition. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 36:914–927, 2014.

[400] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1290–1297. IEEE, 2012.

[401] Jinghua Wang, Zhenhua Wang, Dacheng Tao, Simon See, and Gang Wang. Learning common and specific features for rgb-d semantic segmentation with deconvolutional networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 664–679. Springer, 2016.

[402] Limin Wang, Yu Qiao, and Xiaoou Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 4305–4314. IEEE, 2015.

[403] Limin Wang, Yuanjun Xiong, Zhe Wang, and Yu Qiao. Towards good practices for very deep two-stream convnets. *arXiv preprint arXiv:1507.02159*, 2015.

[404] Pichao Wang, Wanqing Li, Zhimin Gao, Jing Zhang, Chang Tang, and Philip O Ogunbona. Action recognition from depth maps using deep convolutional neural networks. *Transactions on Human-Machine Systems*, 46:498–509, 2016.

[405] Wentao Wang, Wenke Wang, and Sikun Li. From numerics to combinatorics: a survey of topological methods for vector field visualization. *Journal of Visualization*, 19(4):727–752, 2016.

[406] Yang Wang and Greg Mori. Hidden part models for human action recognition: Probabilistic versus max margin. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 33:1310–1323, 2011.

[407] Michael Warren, Luis Mejias, Xilin Yang, Bilal Arain, Felipe Gonzalez, and Ben Upcroft. Enabling aircraft emergency landings using active visual site detection. In *Field and Service Robotics*, pages 167–181. Springer, 2015.

[408] Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. In *Advances in Neural Information Processing Systems*, pages 14334–14345, 2019.

[409] Maurice Weiler, Fred A Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 849–858, 2018.

[410] Matt Whalley, Marc Takahashi, P Tsenkov, G Schulein, and C Goerzen. Field-testing of a helicopter uav obstacle field navigation and landing system. In *65th Annual Forum of the American Helicopter Society, Grapevine, TX*, 2009.

[411] Thomas Whelan, Renato F Salas-Moreno, Ben Glocker, Andrew J Davison, and Stefan Leutenegger. Elasticfusion: Real-time dense slam and light source estimation. *The International Journal of Robotics Research (IJRR)*, 35:1697–1716, 2016.

[412] Geert Willems, Jan Hendrik Becker, Tinne Tuytelaars, and Luc J Van Gool. Exemplar-based action recognition in video. In *Proceedings of the British machine vision conference (BMVC)*, page 3. BMVA Press, 2009.

[413] Geert Willems, Tinne Tuytelaars, and Luc Van Gool. An efficient dense and scale-invariant spatio-temporal interest point detector. In *Proceedings of the European conference on computer vision (ECCV)*, pages 650–663. Springer, 2008.

[414] Paul Wohlhart and Vincent Lepetit. Learning descriptors for object recognition and 3d pose estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3109–3118. IEEE, 2015.

[415] Shu-Fai Wong and Roberto Cipolla. Extracting spatiotemporal interest points using global information. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1–8. IEEE, 2007.

[416] H. Wu, X. Liu, W. An, S. Chen, and H. Lyu. A deep learning approach for efficiently and accurately evaluating the flow field of supercritical airfoils. *Computers & Fluids*, 198:104393, 2020.

[417] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in neural information processing systems 29*, pages 82–90. Curran Associates, Inc., 2016.

[418] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19. Springer, 2018.

[419] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1912–1920. IEEE, 2015.

[420] Lu Xia and JK Aggarwal. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 2834–2841. IEEE, 2013.

[421] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. Sun3d: A database of big spaces reconstructed using sfm and object labels. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1625–1632. IEEE, 2013.

[422] Di Xie, Jiang Xiong, and Shiliang Pu. All you need is beyond a good init: Exploring better solution for training extremely deep convolutional neural networks with orthonormality and modulation. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 6176–6185, 2017.

[423] Huijuan Xu, Kun He, Leonid Sigal, Stan Sclaroff, and Kate Saenko. Text-to-clip video retrieval with early fusion and re-captioning. *arXiv preprint arXiv:1804.05113*, 2018.

[424] Junji Yamato, Jun Ohya, and Kenichiro Ishii. Recognizing human action in time-sequential images using hidden markov model. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 379–385. IEEE, 1992.

[425] Jiaqi Yang, Zhiguo Cao, and Qian Zhang. A fast and robust local descriptor for 3d point cloud registration. *Information Sciences*, 346:163–179, 2016.

[426] Jiaqi Yang, Qian Zhang, Yang Xiao, and Zhiguo Cao. Toldi: An effective and robust approach for 3d local shape description. *Pattern Recognition*, 65:175–187, 2017.

[427] Xiaodong Yang and Ying Li Tian. Eigenjoints-based action recognition using naive-bayes-nearest-neighbor. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 14–19. IEEE, 2012.

[428] Xiaodong Yang and YingLi Tian. Super normal vector for activity recognition using depth sequences. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 804–811. IEEE, 2014.

[429] Lahav Yeffet and Lior Wolf. Local trinary patterns for human action recognition. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 492–497. IEEE, 2009.

[430] Yuichi Yoshida and Takeru Miyato. Spectral norm regularization for improving the generalizability of deep learning. *arXiv preprint arXiv:1705.10941*, 2017.

[431] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *IEEE Computational intelligence magazine*, 13(3):55–75, 2018.

[432] Hongshan Yu, Zhengeng Yang, Lei Tan, Yaonan Wang, Wei Sun, Mingui Sun, and Yandong Tang. Methods and datasets on semantic segmentation: A review. *Neurocomputing*, 304:82–103, 2018.

[433] Tsz-Ho Yu, Tae-Kyun Kim, and Roberto Cipolla. Real-time action recognition by spatiotemporal semantic and structural forests. In *Proceedings of the British machine vision conference (BMVC)*, pages 1–7. BMVA Press, 2010.

[434] Wei Yu, Kuiyuan Yang, Yalong Bai, Hongxun Yao, and Yong Rui. Visualizing and comparing convolutional neural networks. *arXiv preprint arXiv:1412.6631*, 2014.

[435] M Ersin Yumer and Niloy J Mitra. Learning semantic deformation flows with 3d convolutional networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 294–311. Springer, 2016.

[436] Mehmet Ersin Yumer, Siddhartha Chaudhuri, Jessica K Hodgins, and Levent Burak Kara. Semantic shape editing using deformation handles. *ACM Transactions on Graphics (TOG)*, 34(4):1–12, 2015.

[437] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 87.1–87.12, 2016.

[438] Andrei Zaharescu, Edmond Boyer, Kiran Varanasi, and Radu Horaud. Surface feature detection and description with applications to mesh matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 373–380. IEEE, 2009.

[439] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*, 2014.

[440] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 818–833. Springer, 2014.

[441] Jianguo Zhang, Marcin Marszałek, Svetlana Lazebnik, and Cordelia Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision (IJCV)*, pages 213–238, 2007.

[442] Jun-Mei Zhang, Liang Zhong, Boyang Su, Min Wan, Jinq Shya Yap, Jasmine PL Tham, Leok Poh Chua, Dhanjoo N Ghista, and Ru San Tan. Perspective on cfd

studies of coronary artery disease lesions and hemodynamics: A review. *International journal for numerical methods in biomedical engineering*, 30(6):659–680, 2014.

[443] Xin Zhang, Li Liu, Yuxiang Xie, Jie Chen, Lingda Wu, and Matti Pietikainen. Rotation invariant local binary convolution neural networks. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 1210–1219, 2017.

[444] Zhengyou Zhang. Microsoft kinect sensor and its effect. *multimedia*, 19:4–10, 2012.

[445] Bo Zhao, Jiashi Feng, Xiao Wu, and Shuicheng Yan. A survey on deep learning-based fine-grained object classification and semantic segmentation. *International Journal of Automation and Computing*, 14(2):119–135, 2017.

[446] Rui Zhao, Haider Ali, and Patrick Van der Smagt. Two-stream rnn/cnn for action recognition in 3d videos. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4260–4267. IEEE, 2017.

[447] Liang Zheng, Yi Yang, and Qi Tian. Sift meets cnn: A decade survey of instance retrieval. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 40(5):1224–1244, 2017.

[448] Yu Zhong. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *Proceedings of the IEEE international conference on computer vision workshops (ICCVW)*, pages 689–696. IEEE, 2009.

[449] Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. In *Proceedings of the IEEE Conference on computer vision and pattern recognition (CVPR)*, pages 519–528, 2017.

[450] Yu Zou, Xueqian Wang, Tao Zhang, Bin Liang, Jingyan Song, and Houde Liu. Broph: An efficient and compact binary descriptor for 3d point clouds. *Pattern Recognition*, 76:522–536, 2018.