2018/11/30　13:45-14:45

確率場と深層学習に関する第2回CRESTシンポジウム

http://randomfield.cs.waseda.ac.jp/index.php/symposium2

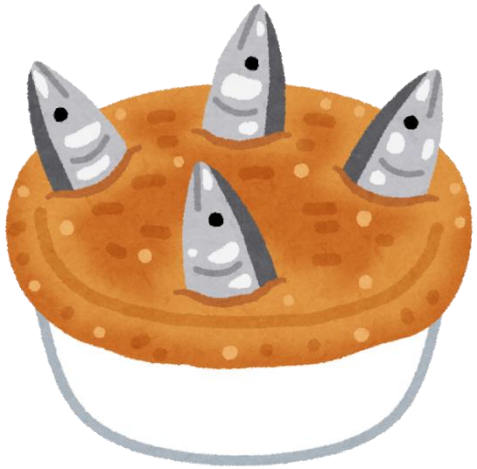
深層学習を用いた三次元物体認識

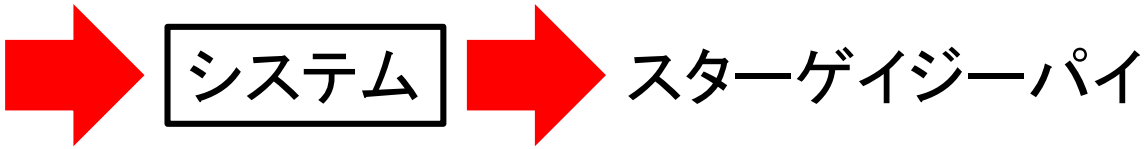
産業技術総合研究所　人工知能研究センター

金崎 朝子

3D物体認識とは

- 3Dデータを入力し、物体のカテゴリ推定結果を出力すること（物体識別）

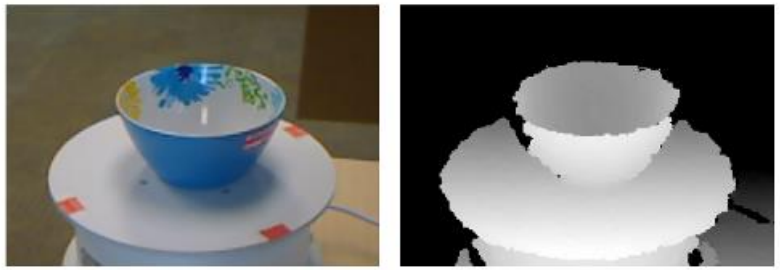




Cf.) 物体検出、物体検索、パーツセグメンテーション

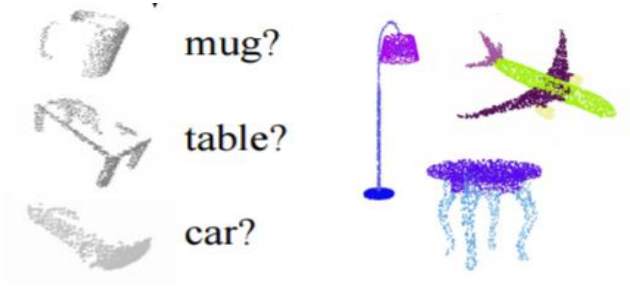
3D物体認識の分類

RGBDベース



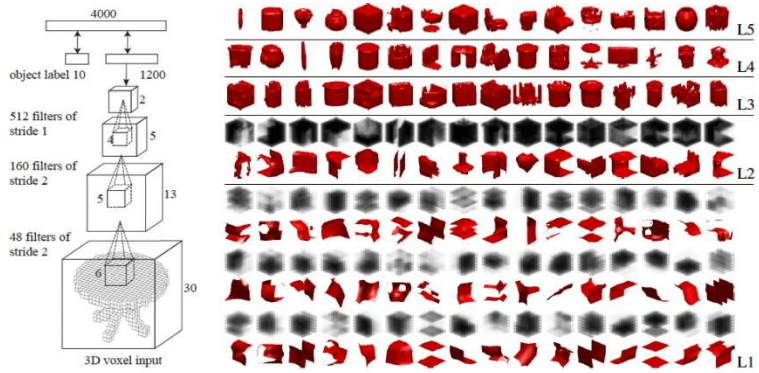
K. Lai et al., **Sparse Distance Learning for Object Recognition Combining RGB and Depth Information.** *ICRA*, 2011.

Point Cloudベース



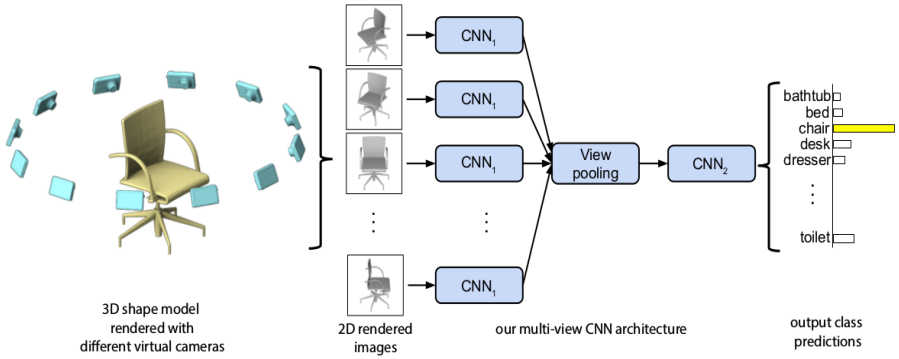
C. Qi et al., **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.** *CVPR*, 2017.

Voxelベース



Z. Wu et al., **3D ShapeNets: A Deep Representation for Volumetric Shape Modeling.** *CVPR*, 2015.

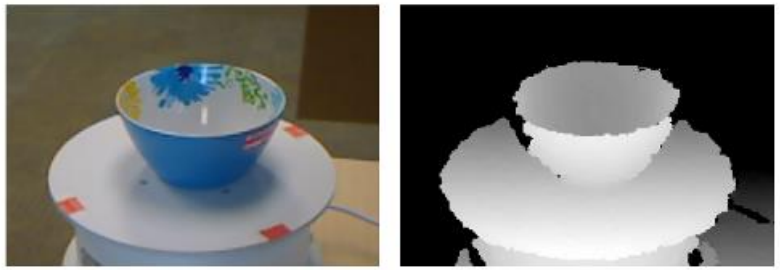
Multi-viewベース



H. Su et al., **Multi-view Convolutional Neural Networks for 3D Shape Recognition.** *ICCV*, 2015.

3D物体認識の分類

RGBDベース



K. Lai et al., **Sparse Distance Learning for Object Recognition Combining RGB and Depth Information.** *ICRA*, 2011.

Point Cloudベース

C. Qi et al., **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.** *CVPR*, 2017.

Voxelベース

Z. Wu et al., **3D ShapeNets: A Deep Representation for Volumetric Shape Modeling.** *CVPR*, 2015.

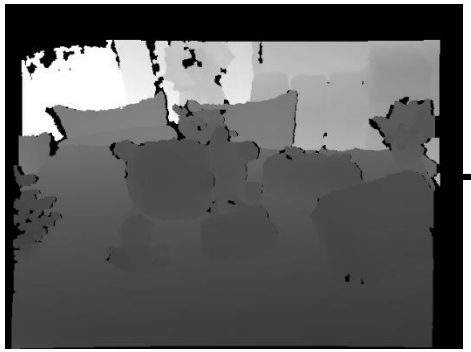
Multi-viewベース

H. Su et al., **Multi-view Convolutional Neural Networks for 3D Shape Recognition.** *ICCV*, 2015.

RGBDベース

RGB画像

D画像



CNN

CNN

認識結果

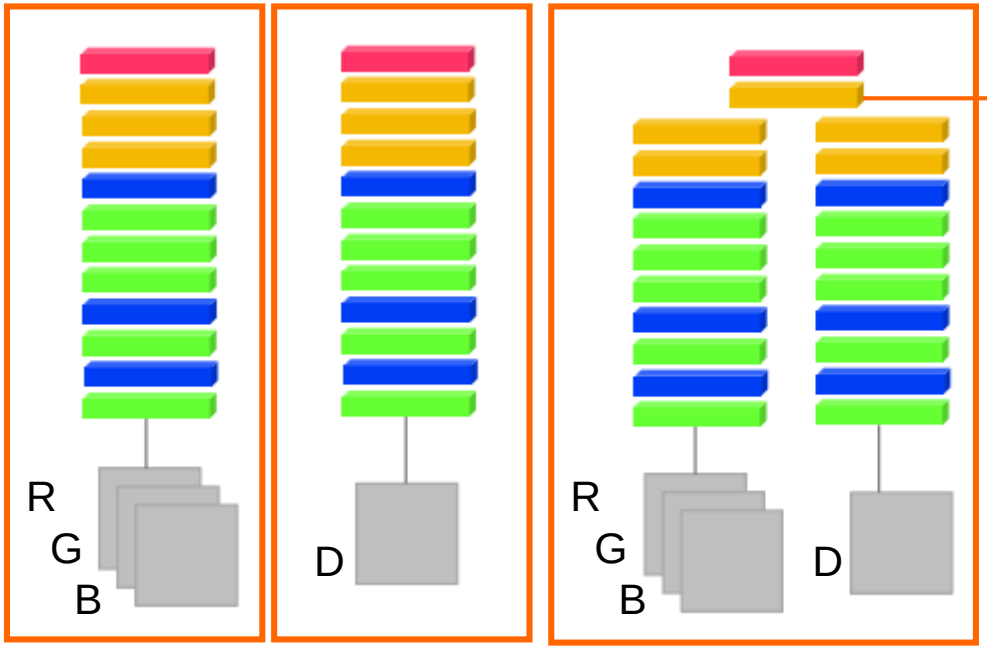
RGBDベースの3D物体認識（1/4）

MMSS: Multi-modal Sharable and Specific Feature Learning for RGB-D Object Recognition

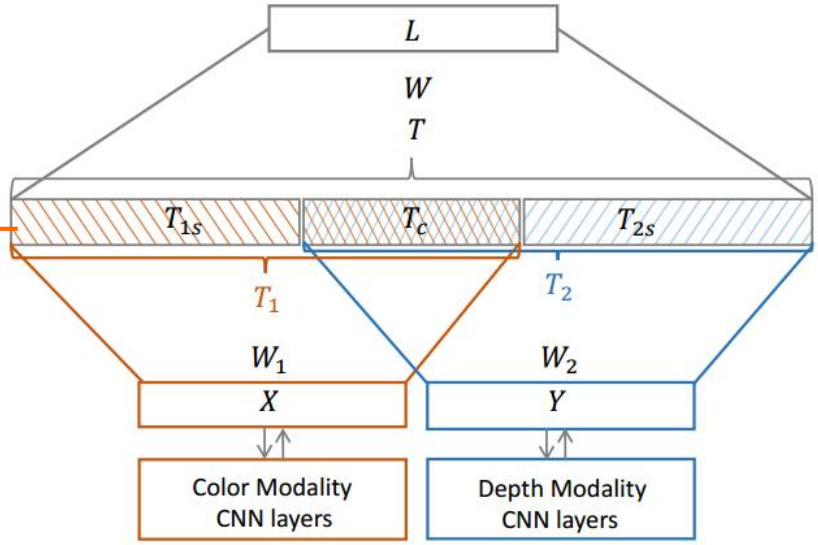
Anran Wang, Jianfei Cai, Jiwen Lu, and Tat-Jen Cham. *IEEE ICCV*, 2015.

pre-training

multi-modal feature learning



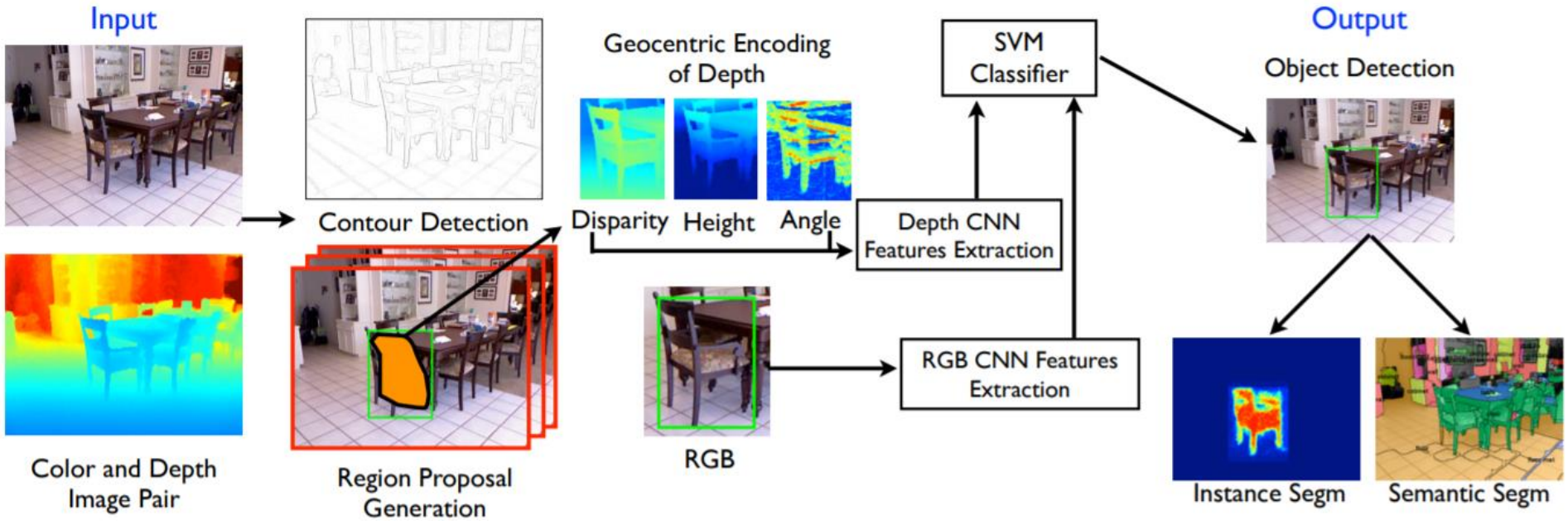
最後のfully-connected層で
RGBとDepthが共通部分を持つよう
Deep CNNを学習する



RGBDベースの3D物体認識（2/4）

Learning Rich Features from RGB-D Images for Object Detection and Segmentation

Saurabh Gupta, Ross Girshick, Pablo Arbelaez, and Jitendra Malik. *ECCV*, 2014.



HHA: **horizontal** disparity, **height** above ground, and **angle** with gravityの3チャンネル

RGBDベースの3D物体認識（3/4）

Depth CNNs for RGB-D scene recognition: learning from scratch better than transferring from RGB-CNNs

Xinhang Song, Luis Herranz, Shuqiang Jiang. *AAAI*, 2017.

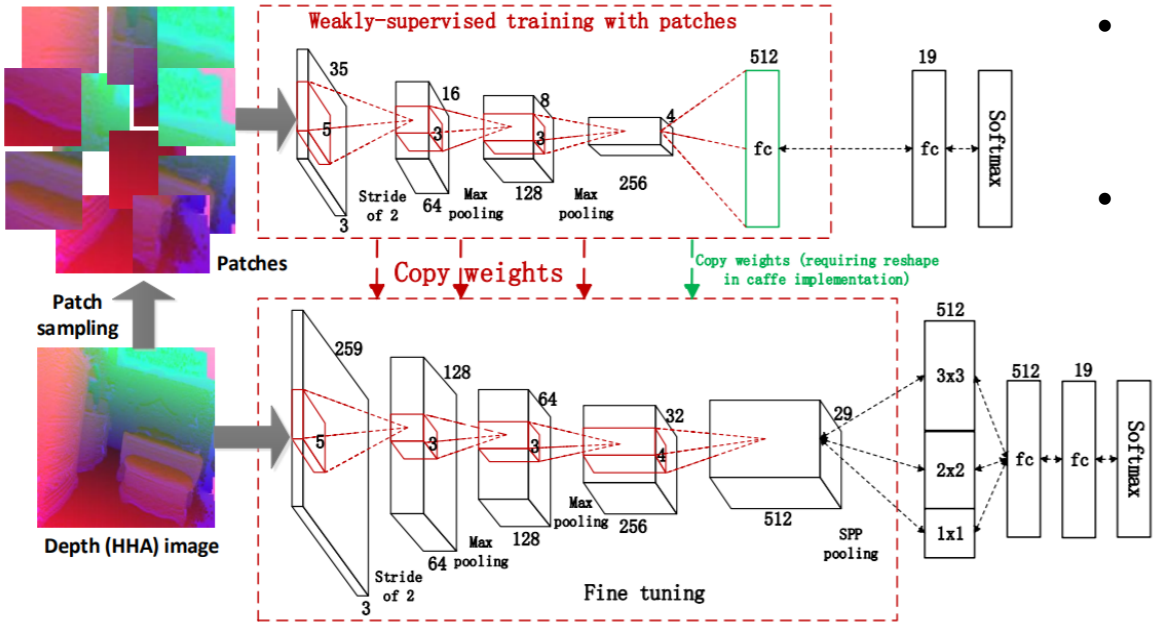


Figure 5: Two-step learning of depth CNNs combining weakly supervised pretraining and fine tuning.

- Depth画像はHHAコーディング、RGB画像のCNNをFine-tuningするのが常套手段。
- Depth CNNをスクラッチから学習する手法の提案。

Table 1: Ablation study for different models (accuracy %).

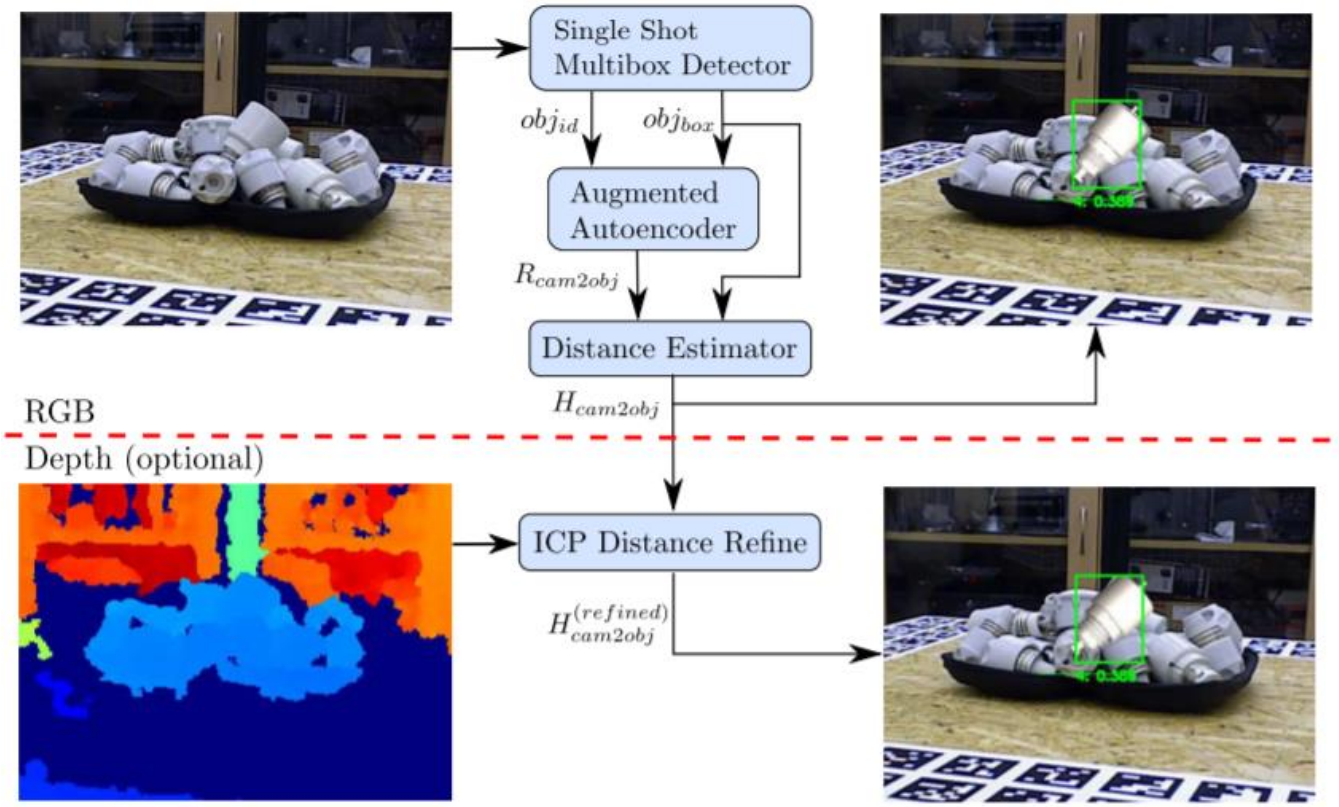
Arch.	Alex-CNN				D-CNN
Weights	Places-CNN		Scratch		Scratch
Layer	-	FT	Train	WSP	WSP
pool1	17.2	20.3	22.3	23.5	**25.3**
pool2	25.3	27.5	26.8	30.4	**33.9**
conv3	27.6	29.3	29.8	**35.1**	34.6
conv4	29.5	32.1	-	-	**38.3**
pool5	30.5	35.9	-	-	-
fc6	30.8	36.5	30.7	36.1	-
fc7	30.9	37.2	32.0	36.8	**40.5**
fc8	-	37.8	32.8	37.5	**41.2**

RGBDベースの3D物体認識（4/4）

Implicit 3D Orientation Learning for 6D Object Detection from RGB Images

M. Sundermeyer, Z. Marton, M. Durner, M. Brucker, and R. Triebel. *ECCV*, 2018.

BEST PAPER AWARD



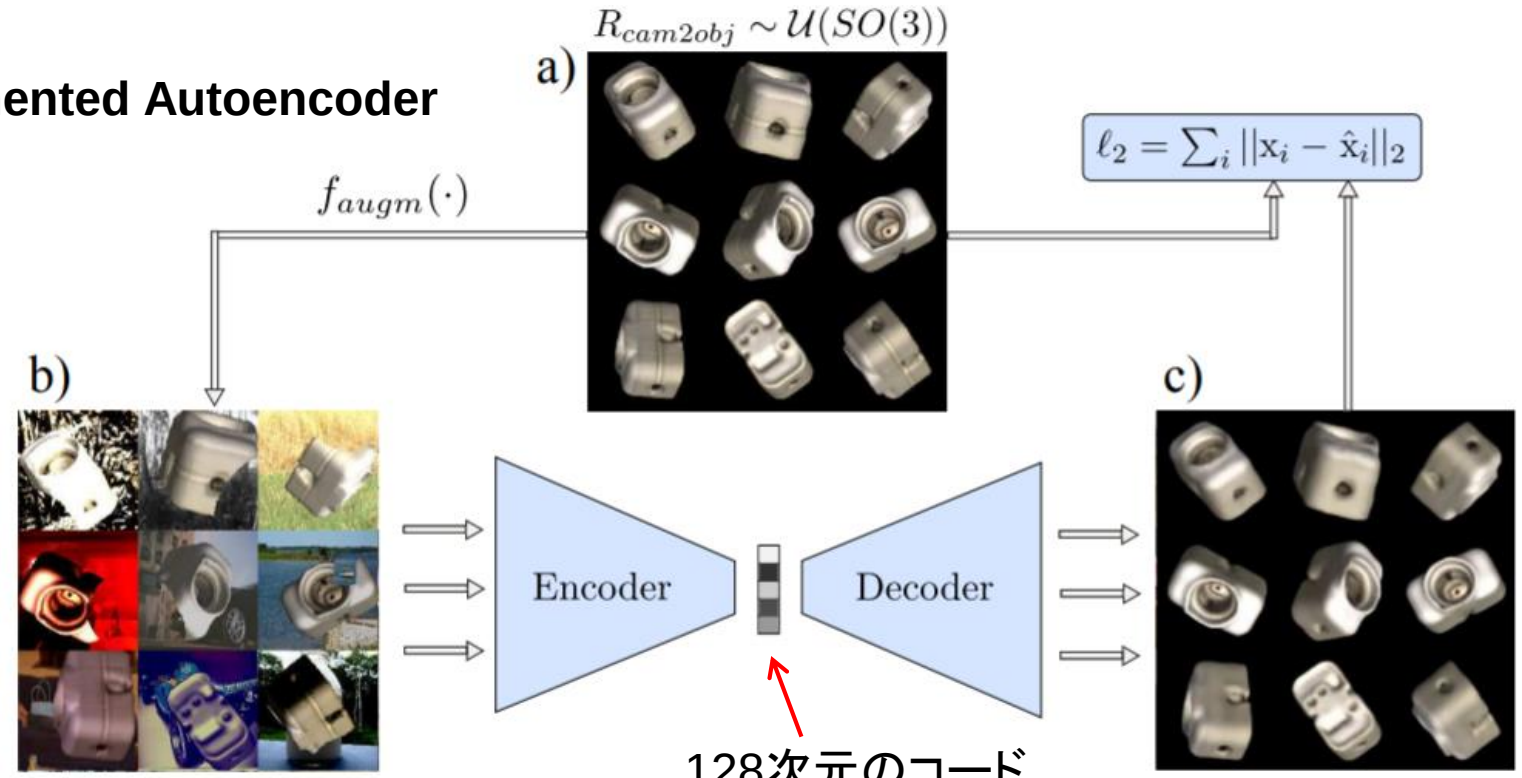
RGBDベースの3D物体認識（4/4）

Implicit 3D Orientation Learning for 6D Object Detection from RGB Images

M. Sundermeyer, Z. Marton, M. Durner, M. Brucker, and R. Triebel. *ECCV*, 2018.

BEST PAPER AWARD

Augmented Autoencoder



128次元のコード
Cosine類似度でリファレンスと比較

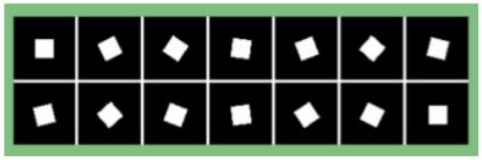
RGBDベースの3D物体認識（4/4）

Implicit 3D Orientation Learning for 6D Object Detection from RGB Images

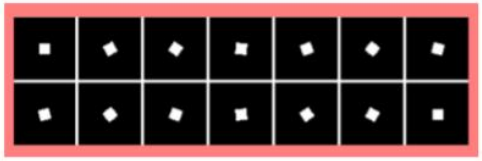
M. Sundermeyer, Z. Marton, M. Durner, M. Brucker, and R. Triebel. *ECCV*, 2018.

BEST PAPER AWARD

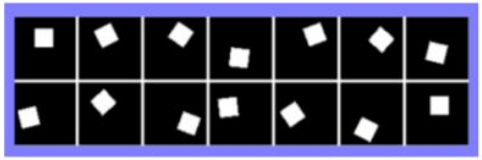
綺麗なの



(a) $X_{s=1.0,t_{xy}=0.0,r\in[0,2\pi]}$

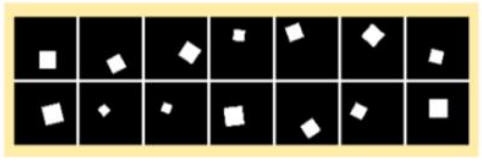


(b) $X_{s=0.6,t_{xy}=0.0,r\in[0,2\pi]}$



(c) $X_{s=1.0,t_{xy}\sim\mathcal{U}(-1,1),r\in[0,2\pi]}$

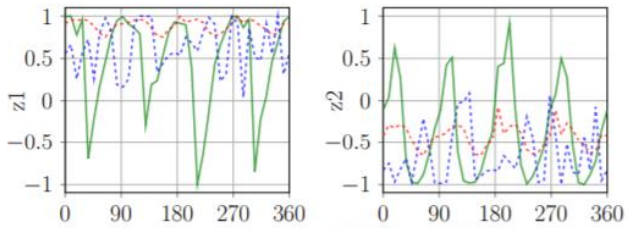
不揃い



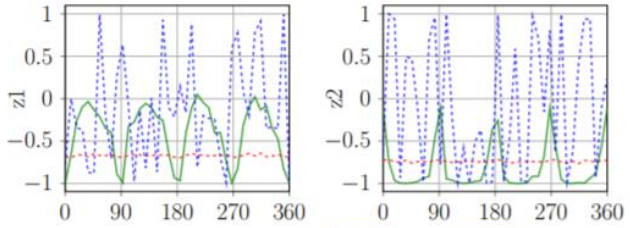
(d) $X_{s\sim\mathcal{U}(0.5,1),t_{xy}\sim\mathcal{U}(-1,1),r\in[0,2\pi]}$

✕綺麗⇒綺麗
✕不揃い⇒不揃い
✔不揃い⇒綺麗

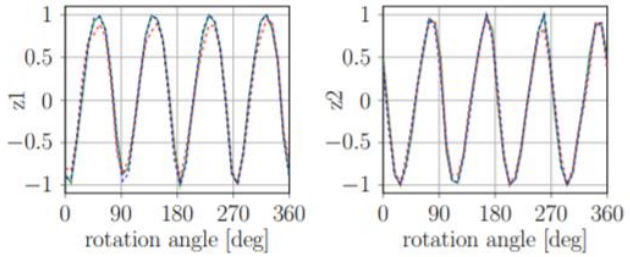
純粋に回転成分を
表す潜在変数を
獲得できる！！



(1) Autoencoder (a) ⟶ (a)



(2) Autoencoder (d) ⟶ (d)



(3) Augmented Autoencoder (d) ⟶ (a)

RGBDベースの3D物体認識（まとめ）

- 基本は2.5次元（1フレームから適用可能）。
- Depth画像はHHAコーディングして、RGB CNNに似たDepth CNNを（Fine-tuning等で）学習するのが一般的。
- 姿勢推定込みの認識によく使われる

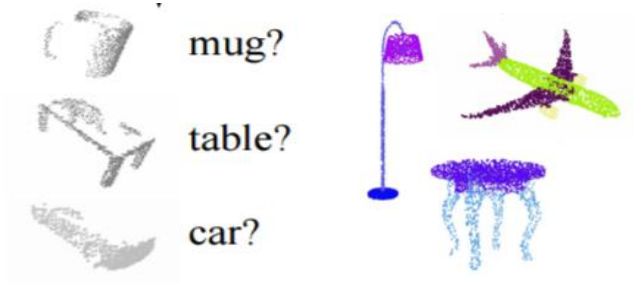
※ただしRGB画像だけでも上手く行っている印象…

3D物体認識の分類

RGBDベース

K. Lai et al., **Sparse Distance Learning for Object Recognition Combining RGB and Depth Information.** *ICRA*, 2011.

Point Cloudベース



C. Qi et al., **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.** *CVPR*, 2017.

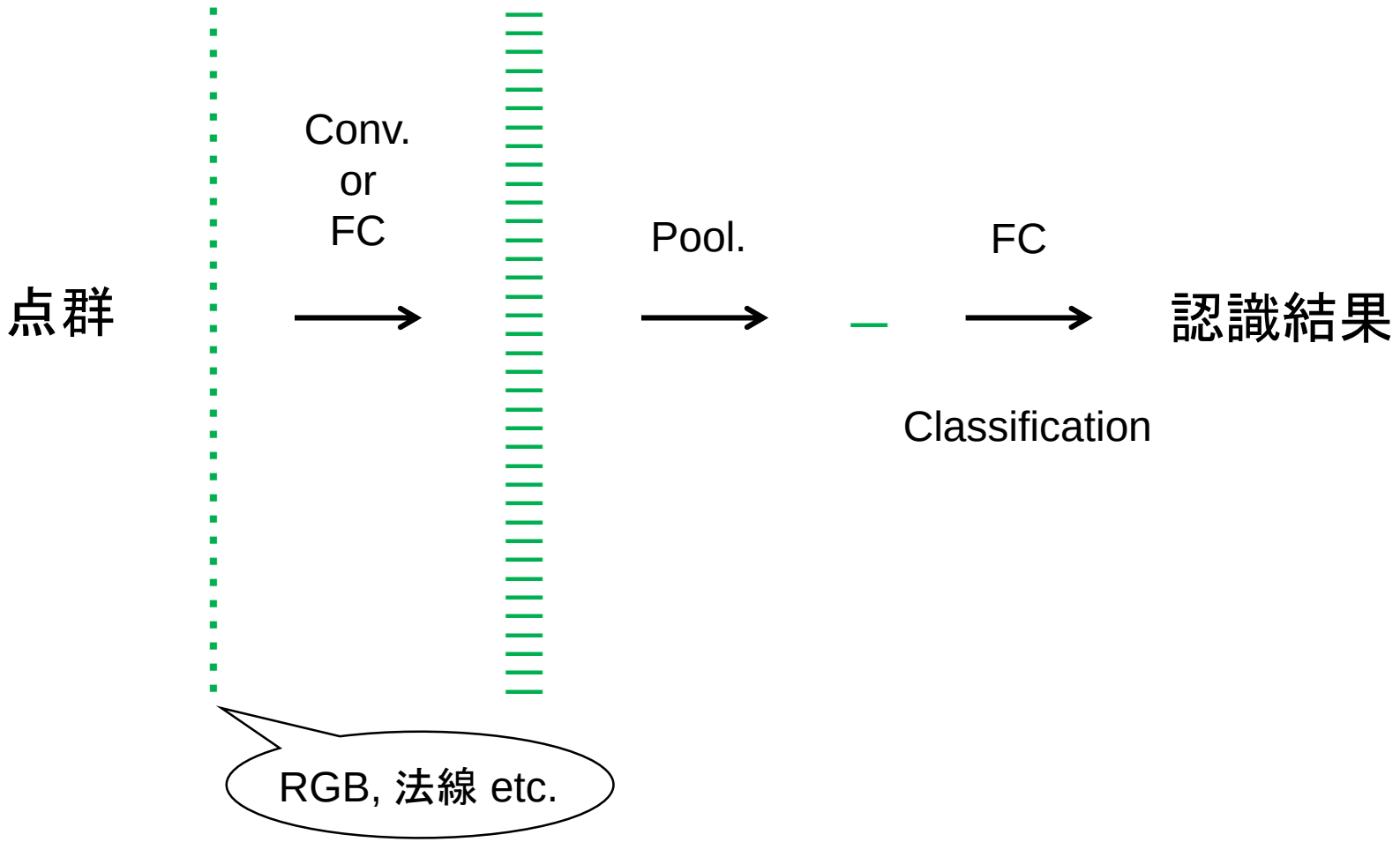
Voxelベース

Z. Wu et al., **3D ShapeNets: A Deep Representation for Volumetric Shape Modeling.** *CVPR*, 2015.

Multi-viewベース

H. Su et al., **Multi-view Convolutional Neural Networks for 3D Shape Recognition.** *ICCV*, 2015.

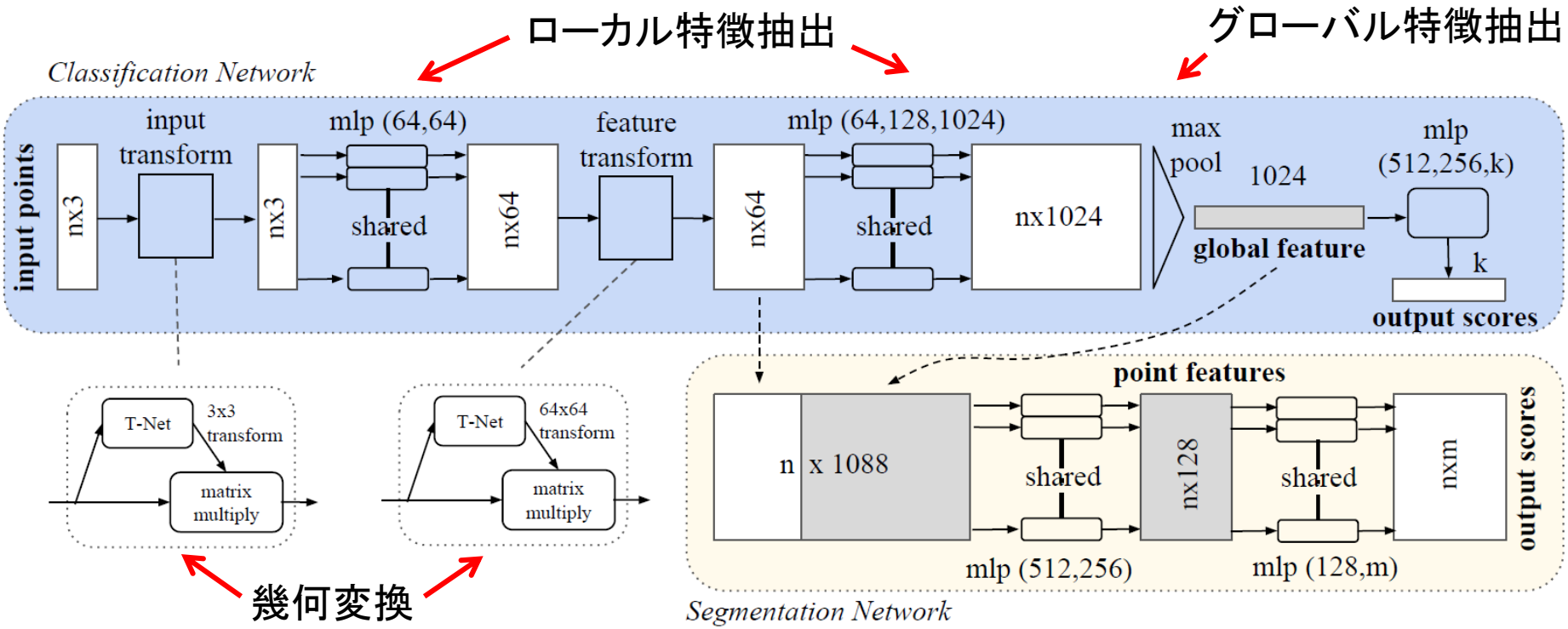
点群ベース



Point Cloudベースの3D物体認識（1/5）

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

Charles R. Qi*, Hao Su*, Kaichun Mo, and Leonidas J. Guibas. *IEEE CVPR*, 2017.

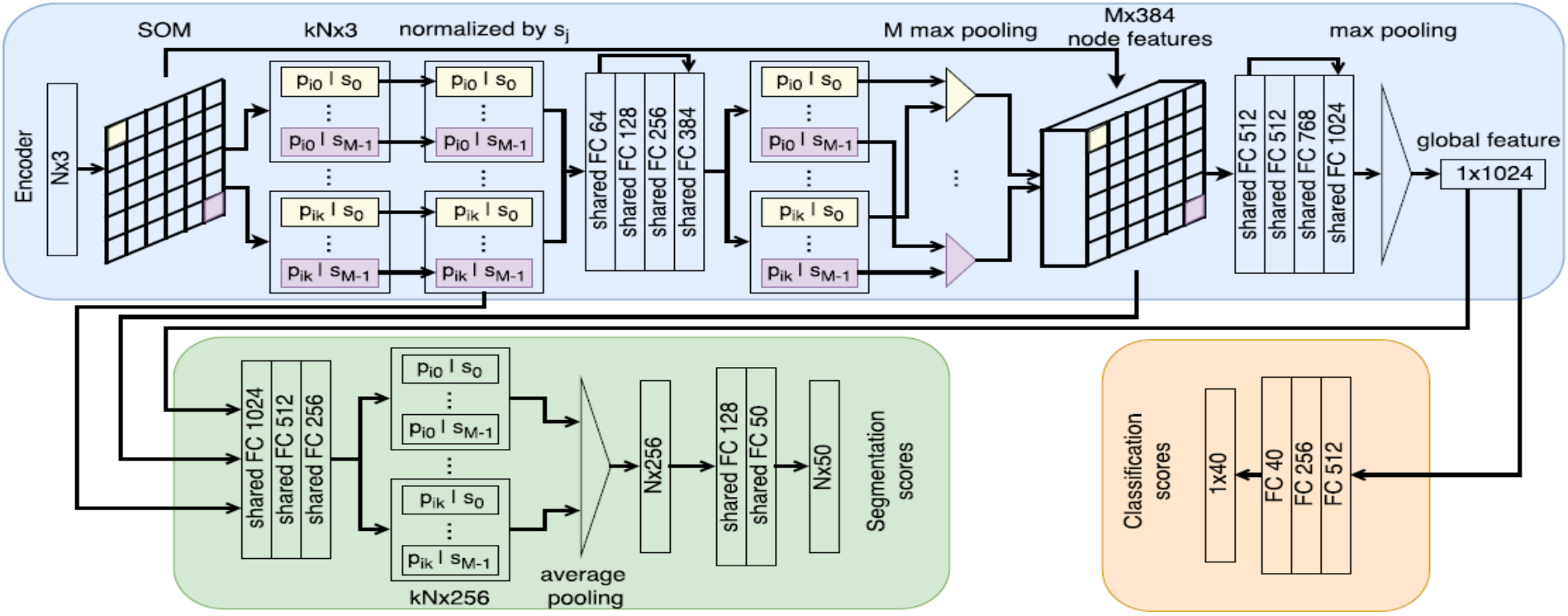


- 回転不変性を確保するため、Sortした点群に直接Multi-layer perceptron（mlp）を適用すると精度が悪い。かわりに、**Max Poolingするのが良かった。**

Point Cloudベースの3D物体認識（2/5）

SO-Net: Self-Organizing Network for Point Cloud Analysis

Jiaxin Li, Ben M. Chen, and Gim Hee Lee. *IEEE CVPR*, 2018.



- 順序不変な自己組織化マップ（SOM）を作り、k近傍探索で点群をSOMノードに割り当てる。点群特徴量はノード毎にMax Pooling→FC層へと渡される。
- 局所特徴量抽出部の雰囲気が従来の点群特徴量に近い。Cf.) FPFH [Rusu et al., 2009]

Point Cloudベースの3D物体認識（2/5）

SO-Net: Self-Organizing Network for Point Cloud Analysis

Jiaxin Li, Ben M. Chen, and Gim Hee Lee. *IEEE CVPR*, 2018.

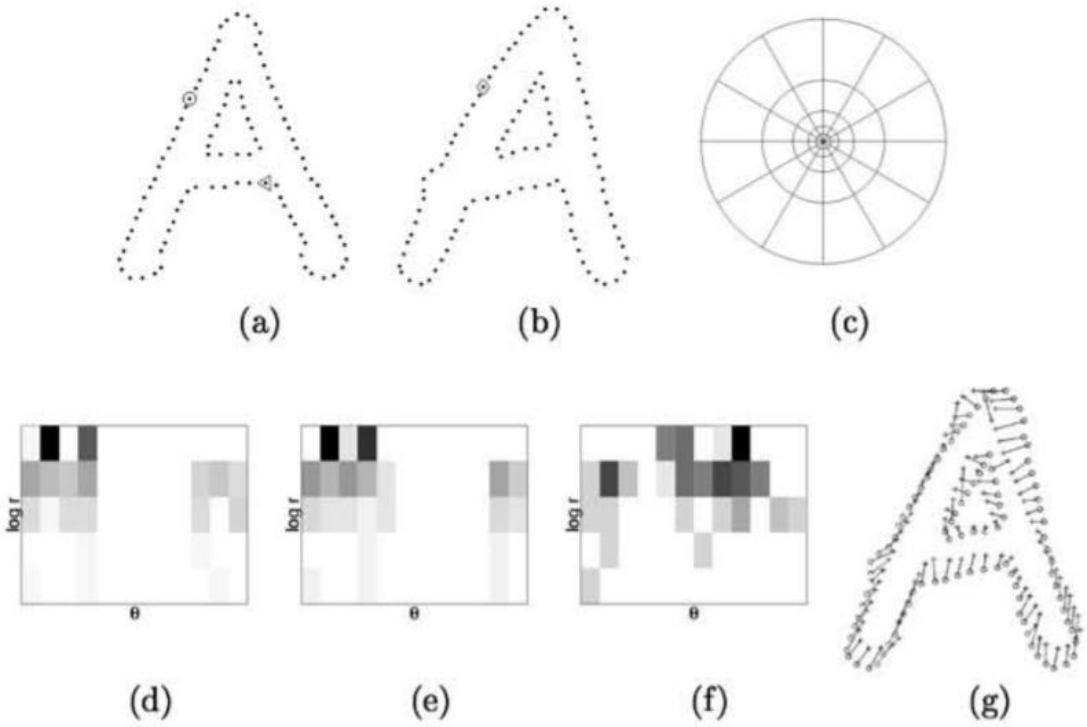
Method	Representation	Input	ModelNet10		ModelNet40			MNIST	
			Class	Instance	Class	Instance	Training	Input	Error rate
PointNet, [26]	points	1024×3	-	-	86.2	89.2	3-6h	256×2	0.78
PointNet++, [28]	points + normal	5000×6	-	-	-	91.9	20h	512×2	0.51
DeepSets, [29, 39]	points	5000×3	-	-	-	90.0	-	-	-
Kd-Net, [18]	points	$2^{15} \times 3$	93.5	94.0	88.5	91.8	120h	1024×2	0.90
ECC, [32]	points	1000×3	90.0	90.8	83.2	87.4	-	-	0.63
OctNet, [30]	octree	128^3	90.1	90.9	83.8	86.5	-	-	-
O-CNN, [36]	octree	64^3	-	-	-	90.6	-	-	-
Ours (2-layer)*	points + normal	5000×6	94.9	95.0	89.4	92.5	3h	-	-
Ours (2-layer)	points + normal	5000×6	94.4	94.5	89.3	92.3	3h	-	-
Ours (2-layer)	points	2048×3	93.9	94.1	87.3	90.9	3h	512×2	**0.44**
Ours (3-layer)	points + normal	5000×6	**95.5**	**95.7**	**90.8**	**93.4**	**3h**	-	-

Table 1. Object classification results for methods using scalable 3D representations like point cloud, kd-tree and octree. Our network produces the best accuracy with significantly faster training speed. * represents pre-training.

- 順序不変な自己組織化マップ（SOM）を作り、k近傍探索で点群をSOMノードに割り当てる。点群特徴量はノード毎にMax Pooling→FC層へと渡される。
- 局所特徴量抽出部の雰囲気が従来の点群特徴量に近い。Cf.) FPFH [Rusu et al., 2009]

従来手法(1)：Shape Context

S. Belongie, J. Malik, and J. Puzicha. "**Shape context: A new descriptor for shape matching and object recognition.**" NIPS, 2001.



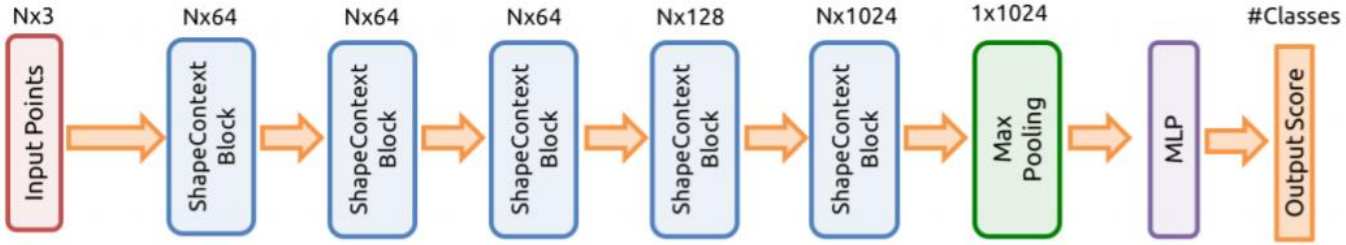
N個の全点につき
他のN-1個の点の
相対座標をビン毎に投票

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in \mathrm{bin}(k)\}$$

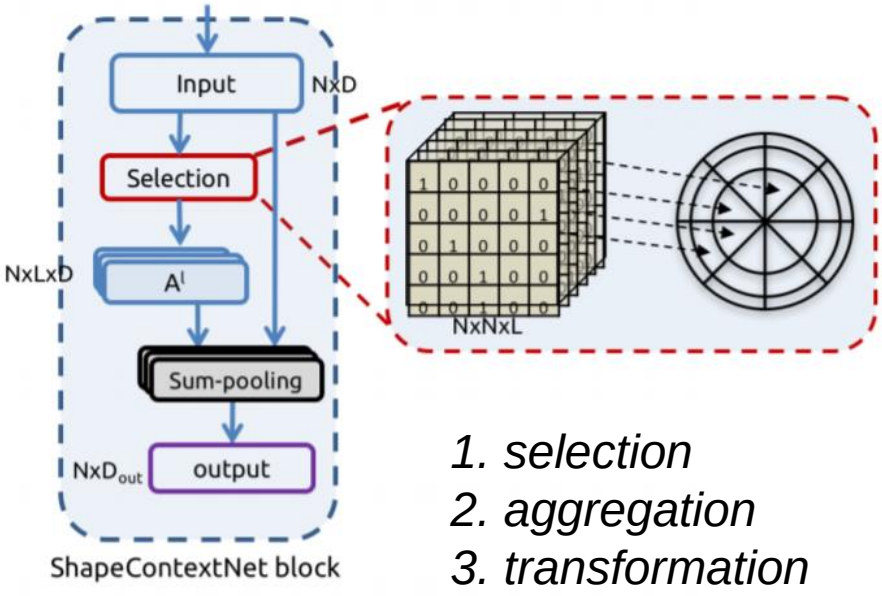
Point Cloudベースの3D物体認識（3/5）

Attentional ShapeContextNet for Point Cloud Recognition

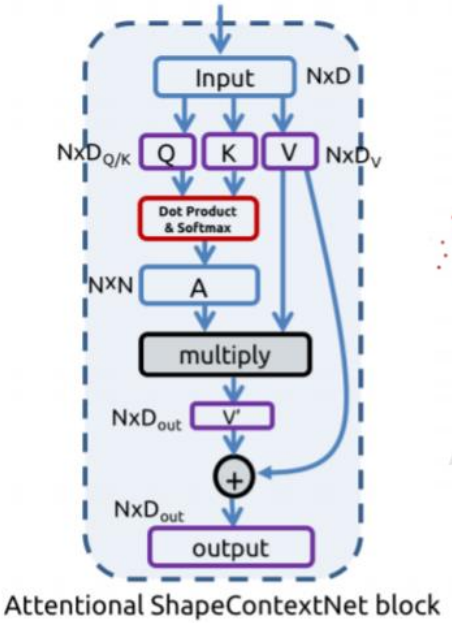
Saining Xie, Sainan Liu, Zeyu Chen, and Zhuowen Tu. CVPR, 2018.

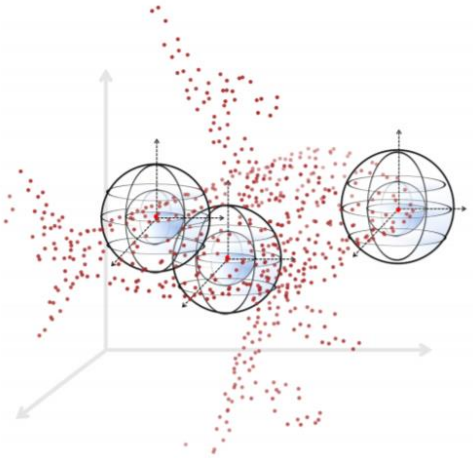


Conv.のかわりに
ShapeContextブロック
NxD⇒NxLxD⇒NxDout



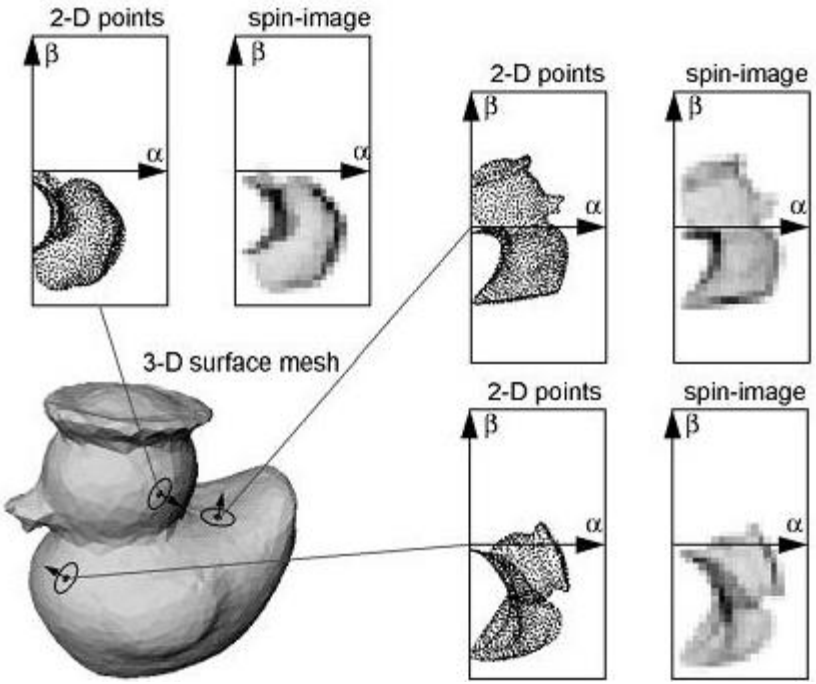
1. selection
2. aggregation
3. transformation

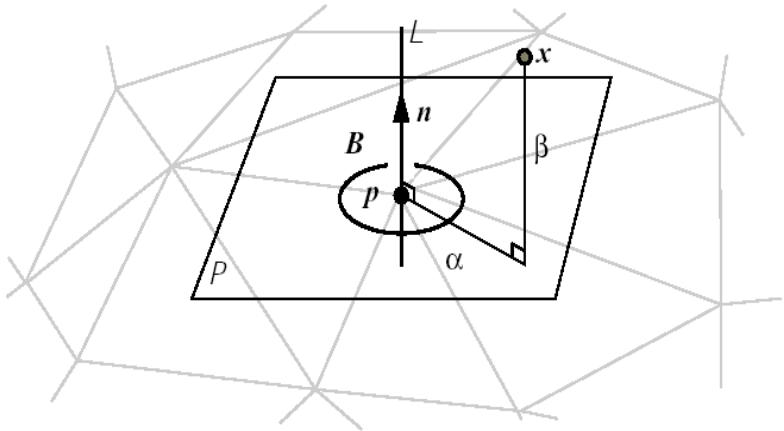




従来手法(2)：Spin Image

A. E. Johnson, and M. Hebert. "**Using spin images for efficient object recognition in cluttered 3D scenes**." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21.5 (1999): 433-449.



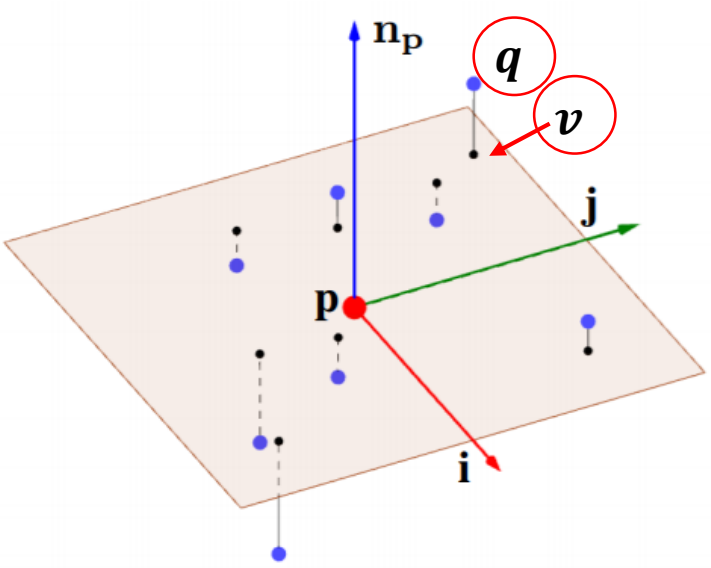


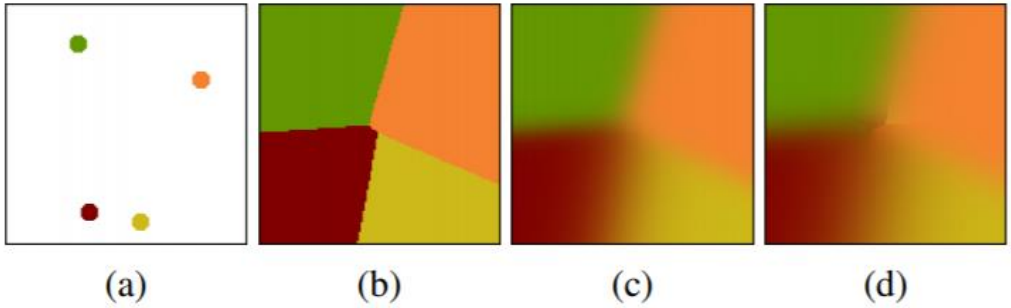
各点のTangent Plane（接平面）に
近傍点を射影し、
射影角αとβの2Dヒストグラムを作る

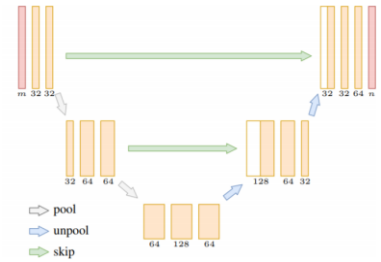
Point Cloudベースの3D物体認識（4/5）

Tangent Convolutions for Dense Prediction in 3D

M. Tatarchenko, J. Park, V. Koltun, Q.-Yi Zhou. CVPR, 2018.







Tangent Convolution:

$$X(\boldsymbol{p}) = \int_{\pi_p} c(\boldsymbol{u}) S(\boldsymbol{u}) d\boldsymbol{u}$$

$$= \int_{\pi_p} c(\boldsymbol{u}) \cdot \sum_{\boldsymbol{v}} (w(\boldsymbol{u}, \boldsymbol{v}) \cdot F(\boldsymbol{q}))\, d\boldsymbol{u}$$

Convolutionカーネル

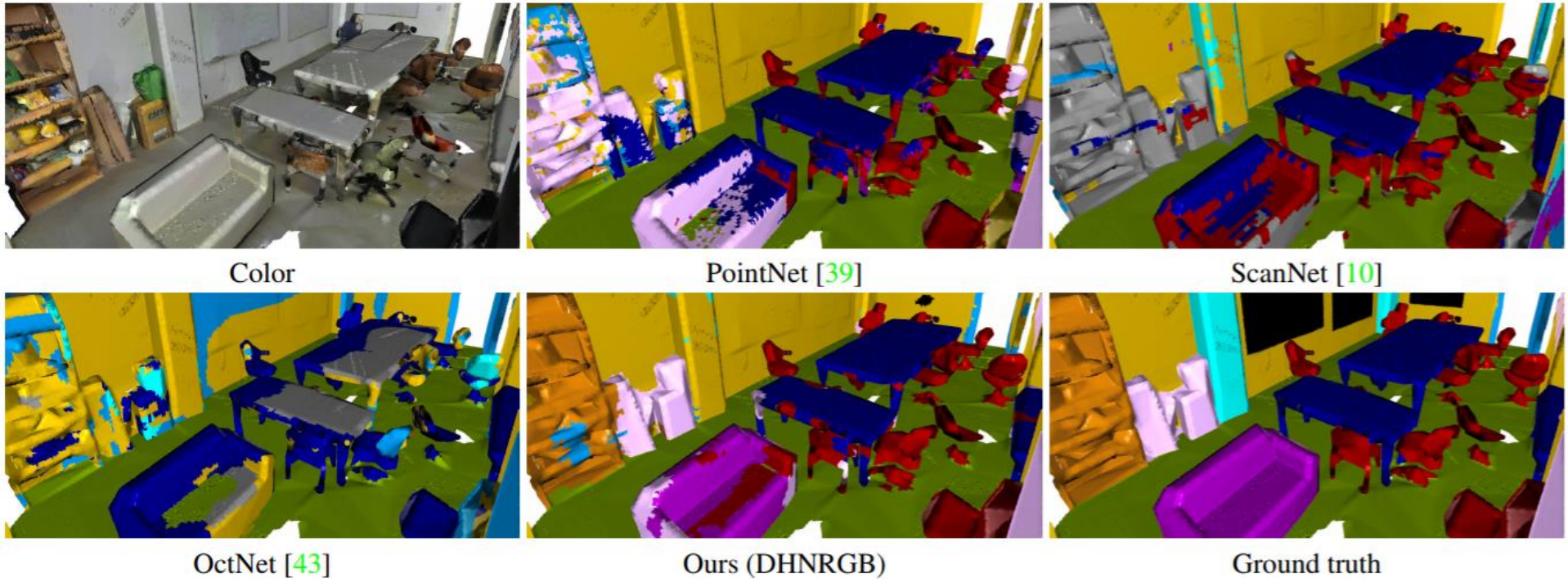
全近傍点$\boldsymbol{q}$の持つ値から補完した値

Point Cloudベースの3D物体認識（4/5）

Tangent Convolutions for Dense Prediction in 3D

M. Tatarchenko, J. Park, V. Koltun, Q.-Yi Zhou. CVPR, 2018.

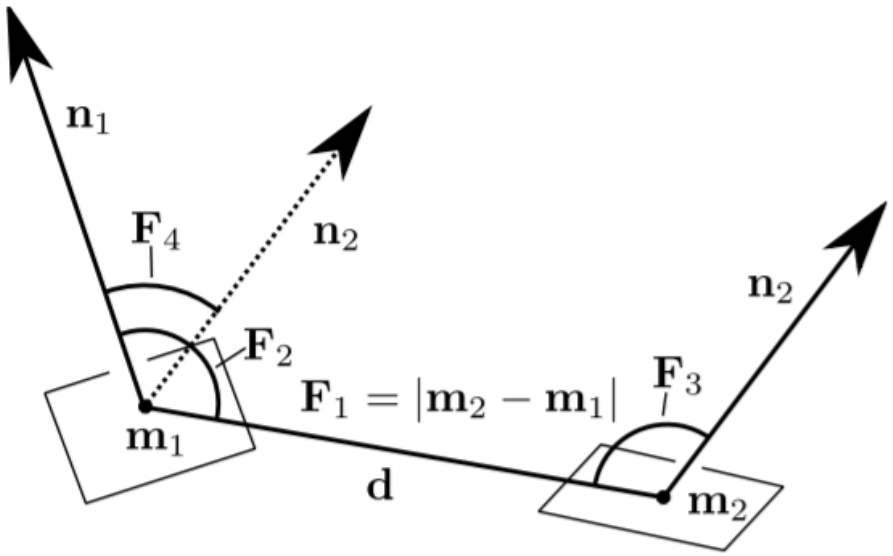
アプリケーションはセマンティックセグメンテーション



従来手法(3)：Point Pair Features (PPF)

Model Globally, Match Locally: Efficient and Robust 3D Object Recognition

B. Drost, M. Ulrich, N. Navab, S. Ilic. CVPR, 2010.



二点間の距離と法線の相対角度で記述される4次元の特徴量

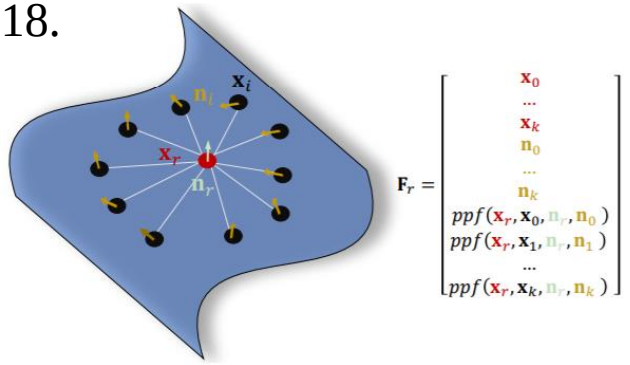
$$\angle(\mathbf{v_1}, \mathbf{v_2}) = \tan^{-1}\left(\frac{\|\mathbf{v_1} \times \mathbf{v_2}\|}{\mathbf{v_1} \cdot \mathbf{v_2}}\right)$$

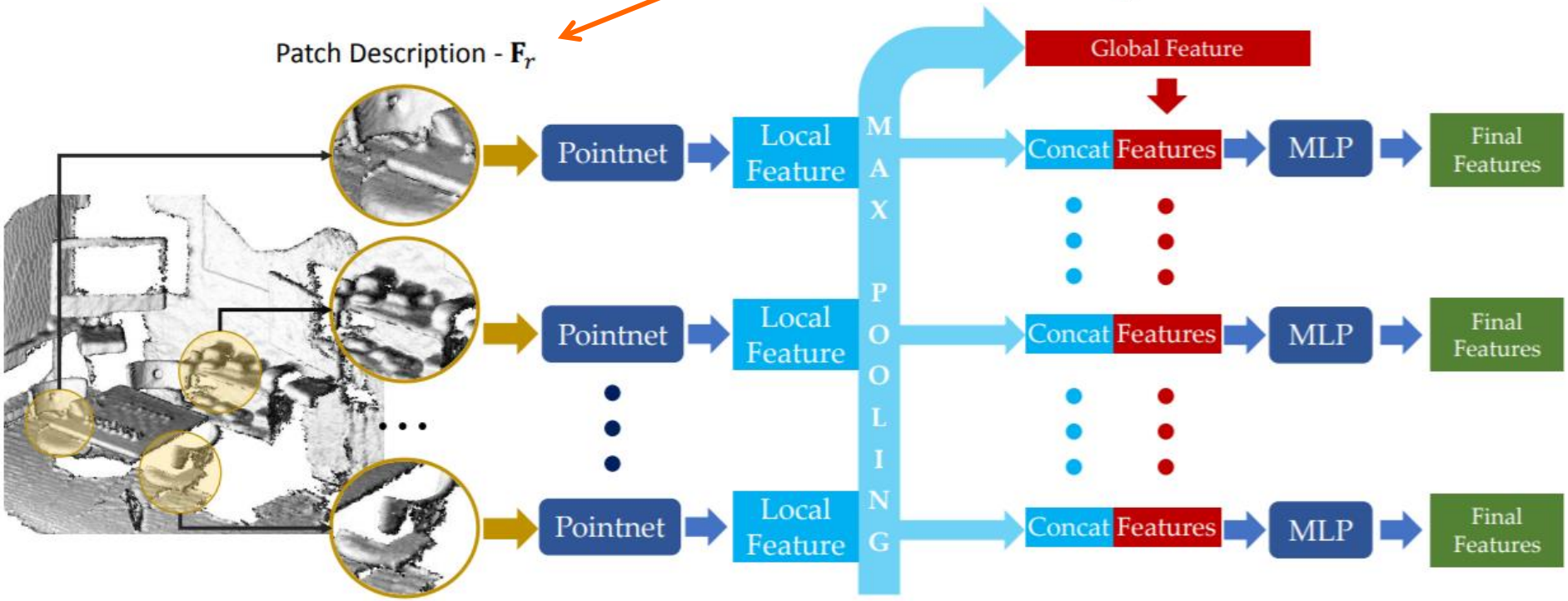
$$\mathbf{F}(\mathbf{m_1}, \mathbf{m_2}) = (\|\mathbf{d}\|_2, \angle(\mathbf{n_1}, \mathbf{d}), \angle(\mathbf{n_2}, \mathbf{d}), \angle(\mathbf{n_1}, \mathbf{n_2}))$$

Point Cloudベースの3D物体認識（5/5）

PPFNet: Global Context Aware Local Features for Robust 3D Point Matching H. Deng, T. Birdal, S. Ilic. CVPR, 2018.

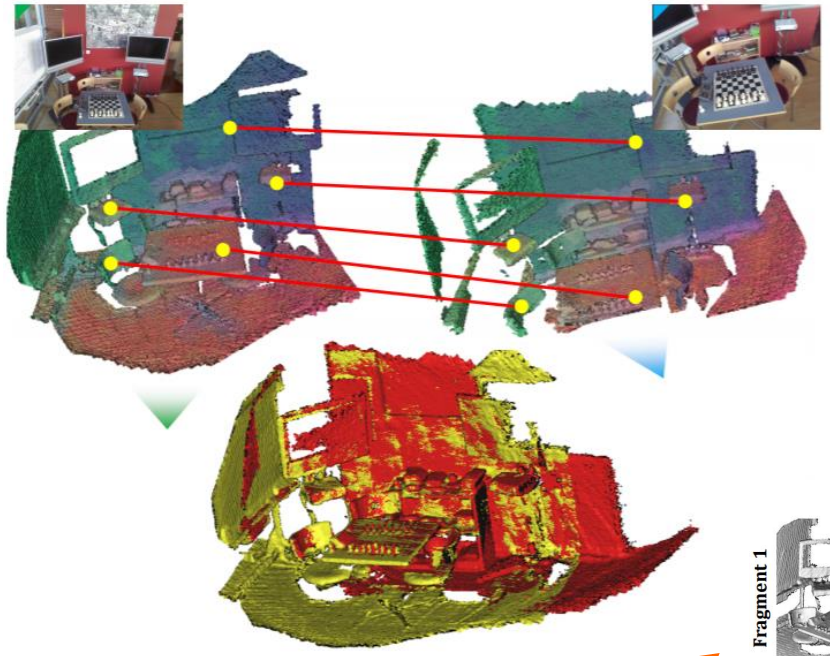
ローカルパッチ毎に点の座標、法線、
PPFを並べたセットを特徴量として
PointNetに入力する





Point Cloudベースの3D物体認識（5/5）

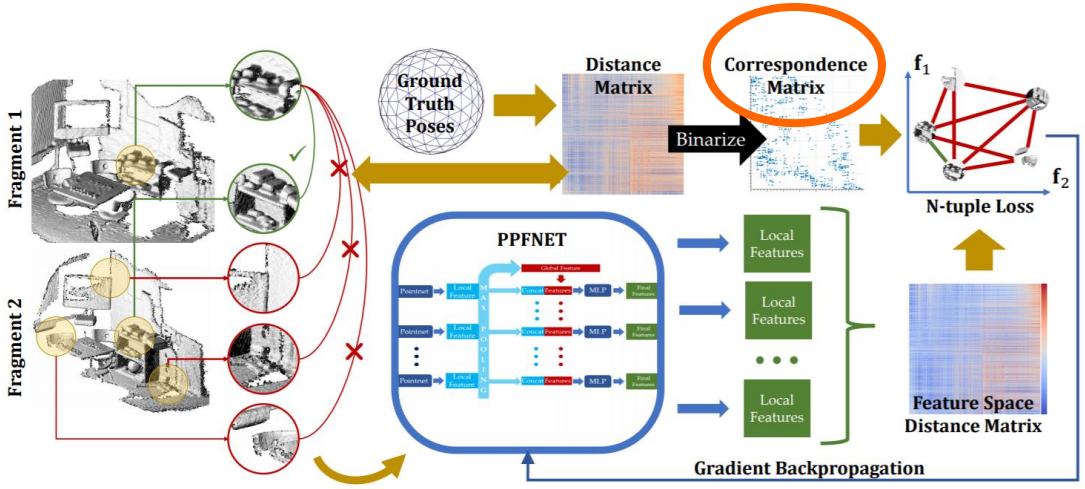
PPFNet: Global Context Aware Local Features for Robust 3D Point Matching H. Deng, T. Birdal, S. Ilic. CVPR, 2018.



アプリケーションは
点群レジストレーション

フラグメントのペアを入力すると
（ローカルパッチの）対応を出力する

Correspondence Matrix



フラグメント1

フラグメント2

参考資料@SlideShare

- “CVPR2018のPointCloudのCNN論文とSPLATNet” – by Takuya Minagawa

https://www.slideshare.net/takmin/cvpr2018pointcloudcnnsplatnet

- “三次元点群を取り扱うニューラルネットワークのサーベイ” – by Naoya Chiba

https://www.slideshare.net/naoyachiba18/ss-120302579

Point Cloudベースの3D物体認識（まとめ）

- 回転不変な局所（ローカル）特徴量をどうとるか。
- 局所（ローカル）特徴量をどう大域（グローバル）特徴量に統合するか。
- 物体の回転に強い。
- パーツセグメンテーションに応用しやすい。

3D物体認識の分類

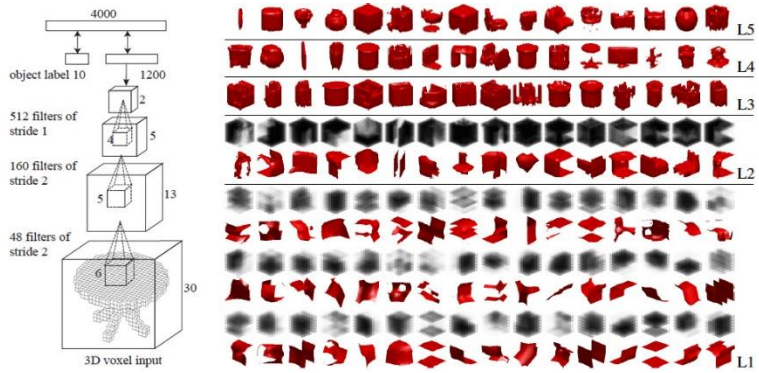
RGBDベース

K. Lai et al., **Sparse Distance Learning for Object Recognition Combining RGB and Depth Information.** *ICRA*, 2011.

Point Cloudベース

C. Qi et al., **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.** *CVPR*, 2017.

Voxelベース

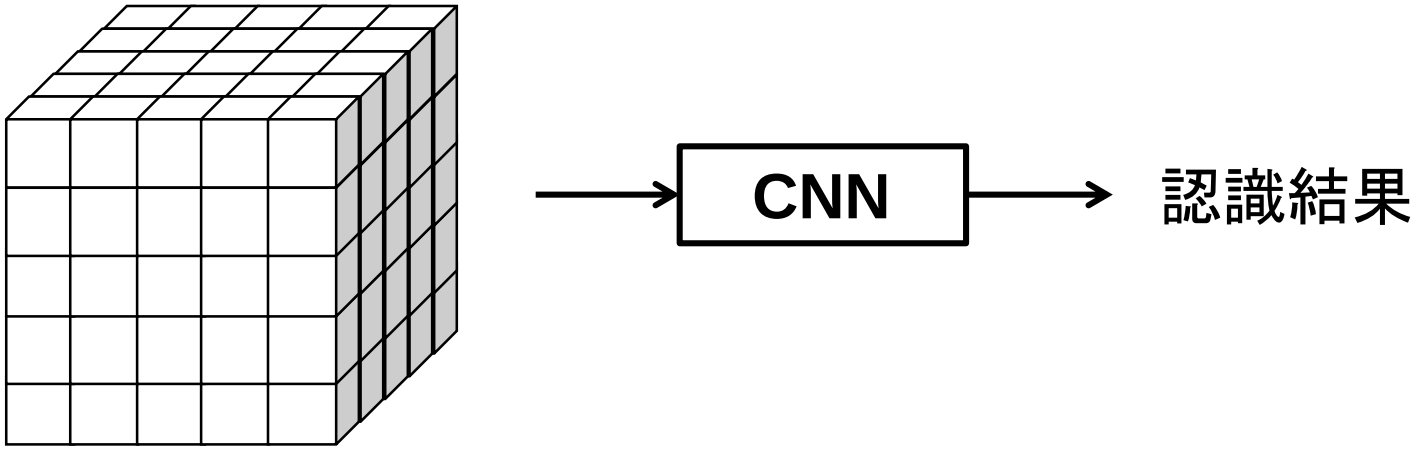


Z. Wu et al., **3D ShapeNets: A Deep Representation for Volumetric Shape Modeling.** *CVPR*, 2015.

Multi-viewベース

H. Su et al., **Multi-view Convolutional Neural Networks for 3D Shape Recognition.** *ICCV*, 2015.

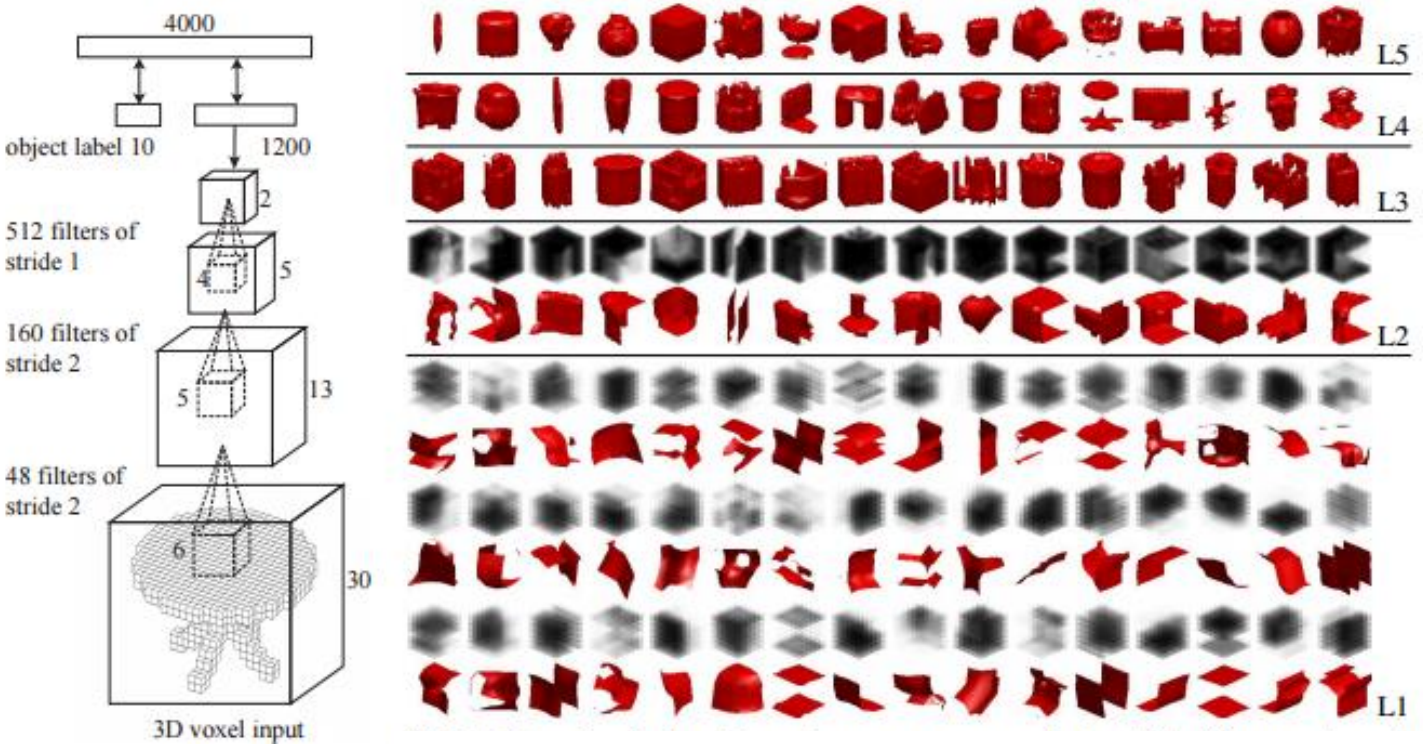
Voxelベース



Voxelベースの3D物体認識（1/4）

3D ShapeNets: A Deep Representation for Volumetric Shapes

Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. *IEEE CVPR,* 2015.

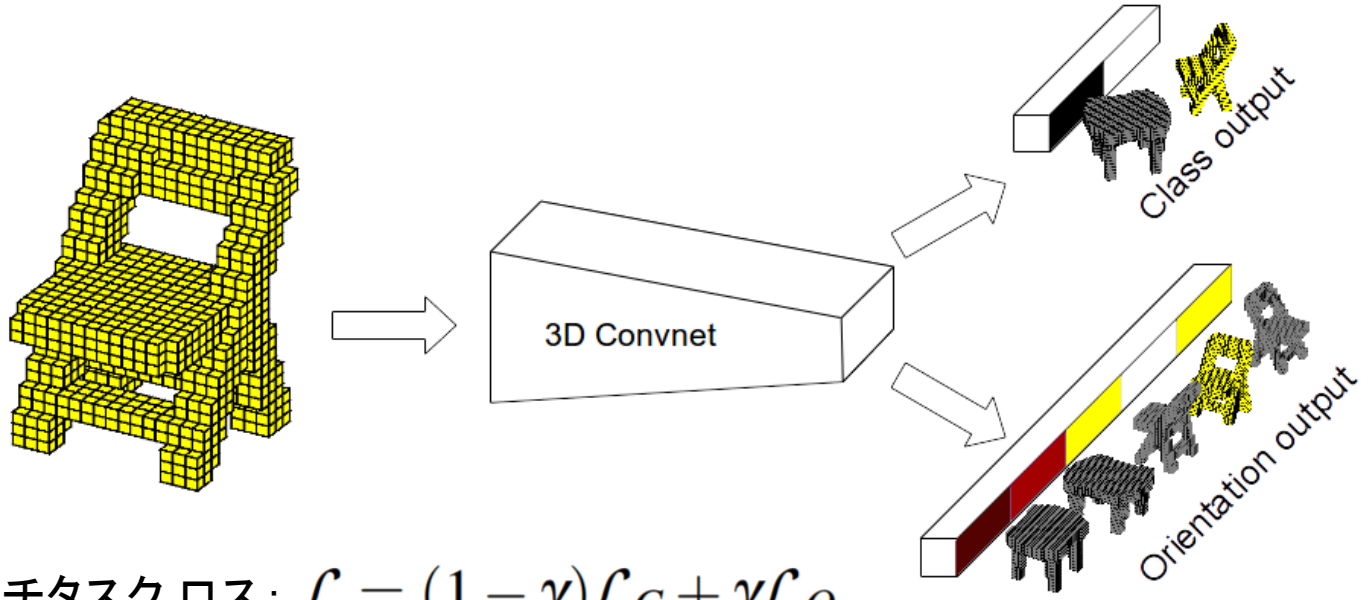


- 151,128 3D CAD models belonging to 660 unique object categories を
- 30 x 30 x 30のボクセルデータに変換して、Deep Learningで学習。
- Light Field descriptor [Chen et al. 2003], Spherical Harmonic descriptor [Kazhdan et al. 2003] と比較して高性能。

Voxelベースの3D物体認識（2/4）

Orientation-boosted Voxel Nets for 3D Object Recognition

N. Sedaghat, M. Zolfaghari, E. Amiri, and T. Brox. *BMVC*, 2017.



$= N$次元

$= N \cdot M$次元

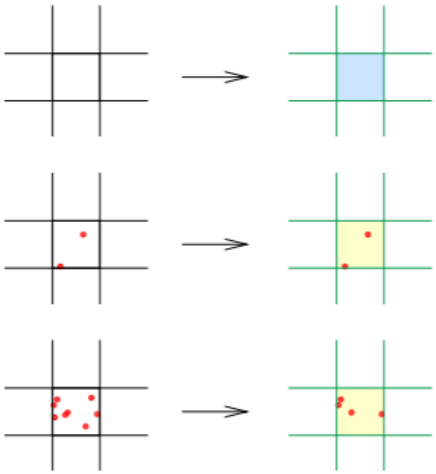
マルチタスク ロス: $\mathcal{L} = (1-\gamma)\mathcal{L}_C + \gamma\mathcal{L}_O$

- 垂直軸は固定で、そのまわり（azimuth）の回転を考える。
- 物体カテゴリ識別と姿勢（オリエンテーション）識別のマルチタスク学習。
- テスト時は複数の回転姿勢のボクセルを入力し、カテゴリスコアを平均する。
- テスト時にOrientation推定は使わない。（！）
- **マルチタスク学習によってカテゴリ識別精度が向上することを示した。**

Voxelベースの3D物体認識（3/4）

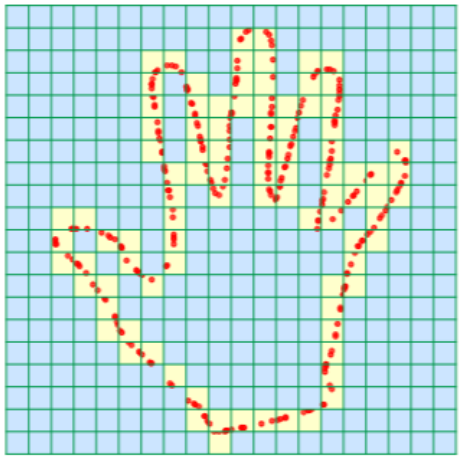
PointGrid: A Deep Network for 3D Shape Understanding

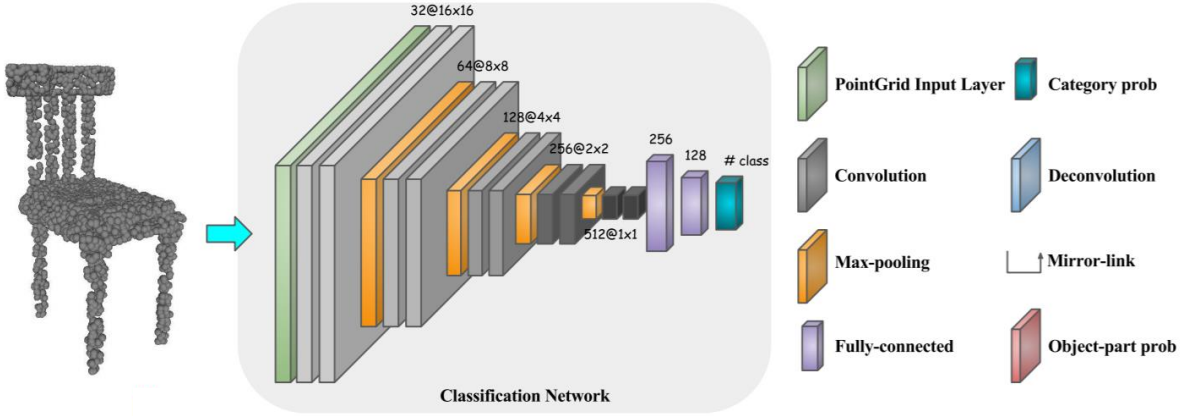
Truc Le and Ye Duan, *CVPR*, 2018.



- 各ボクセルが0個、あるいはK個（一定数）の点を持つようなリサンプリングを行う
- 各ボクセルはK個の点の(x, y, z)座標を連結した$3K$次元の特徴量を持つ

ボクセル解像度が粗い問題を解決！





Classification Network

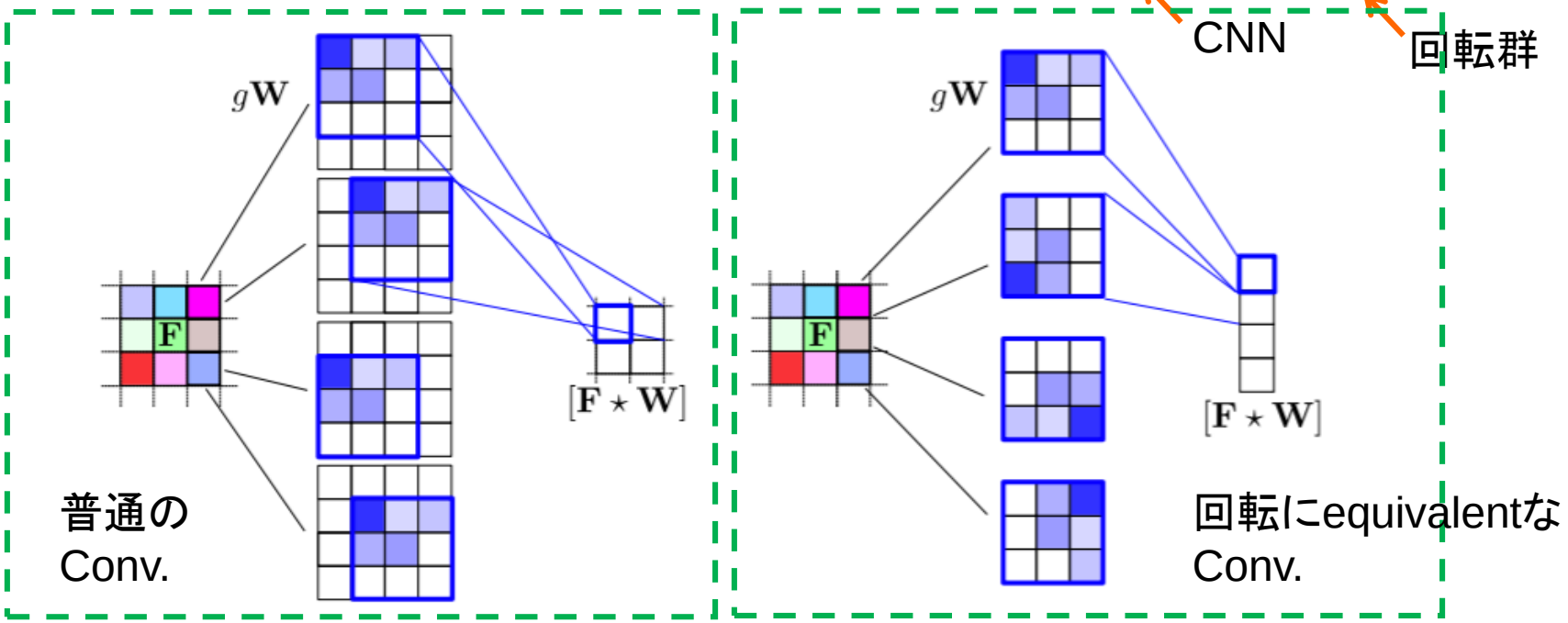
Voxelベースの3D物体認識（4/4）

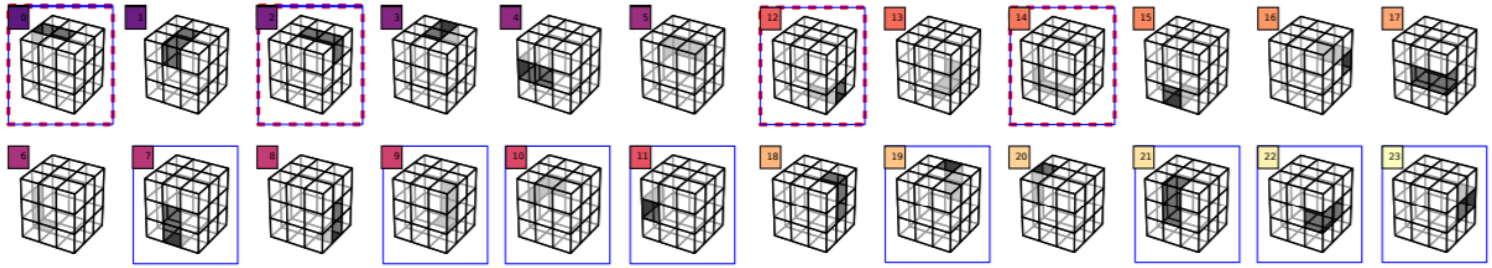
CubeNet: Equivariance to 3D Rotation and Translation

Daniel Worrall and Gabriel Brostow, *ECCV*, 2018.

$$\Phi(g\mathbf{x}) = g\Phi(\mathbf{x})$$

CNN　回転群





Cube Group

Voxelベースの3D物体認識（まとめ）

- 低解像度（にせざるを得ない）のため認識精度は高くない。
  - VoxelGridのような工夫が必要
  - （アーキテクチャを改良すれば精度は上がるような気がする。）
- 回転にどう対応するか？という問題がある。
  - CubeNetのようなのがあるがサンプリングが回転依存な問題は解消されていない

3D物体認識の分類

RGBDベース

K. Lai et al., **Sparse Distance Learning for Object Recognition Combining RGB and Depth Information.** *ICRA*, 2011.

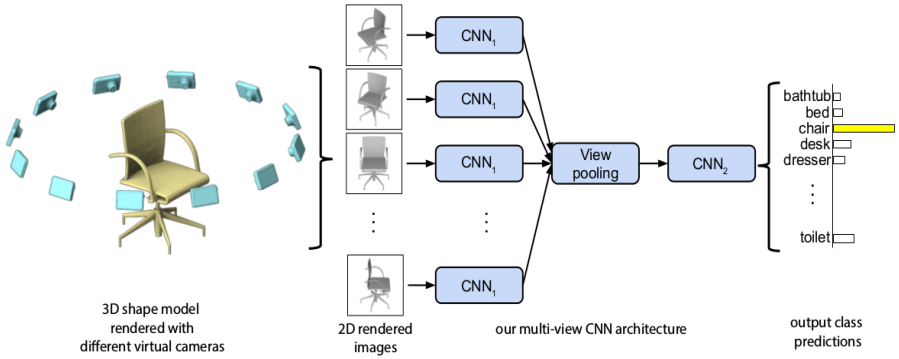
Point Cloudベース

C. Qi et al., **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.** *CVPR*, 2017.

Voxelベース

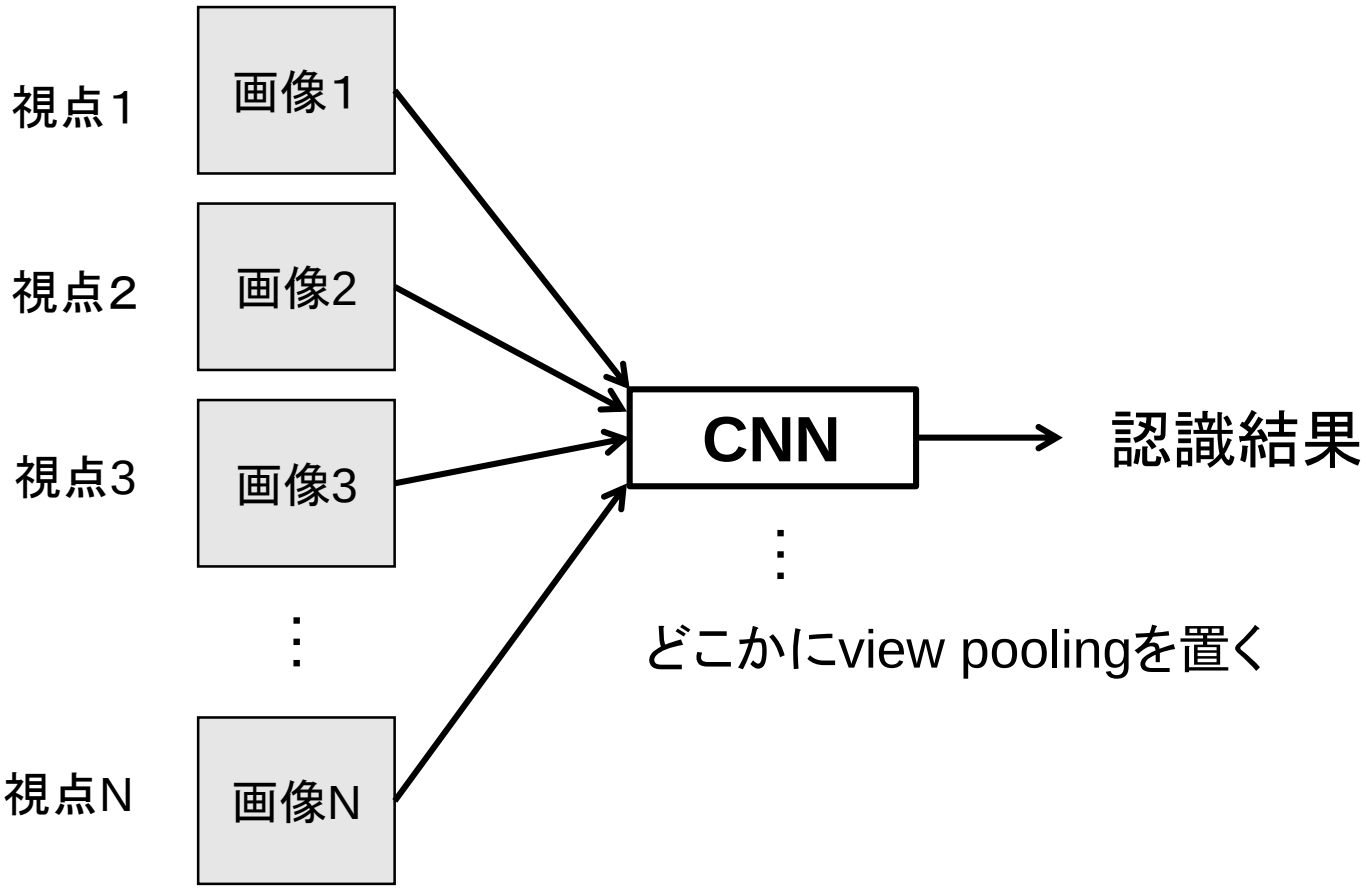
Z. Wu et al., **3D ShapeNets: A Deep Representation for Volumetric Shape Modeling.** *CVPR*, 2015.

Multi-viewベース



H. Su et al., **Multi-view Convolutional Neural Networks for 3D Shape Recognition.** *ICCV*, 2015.

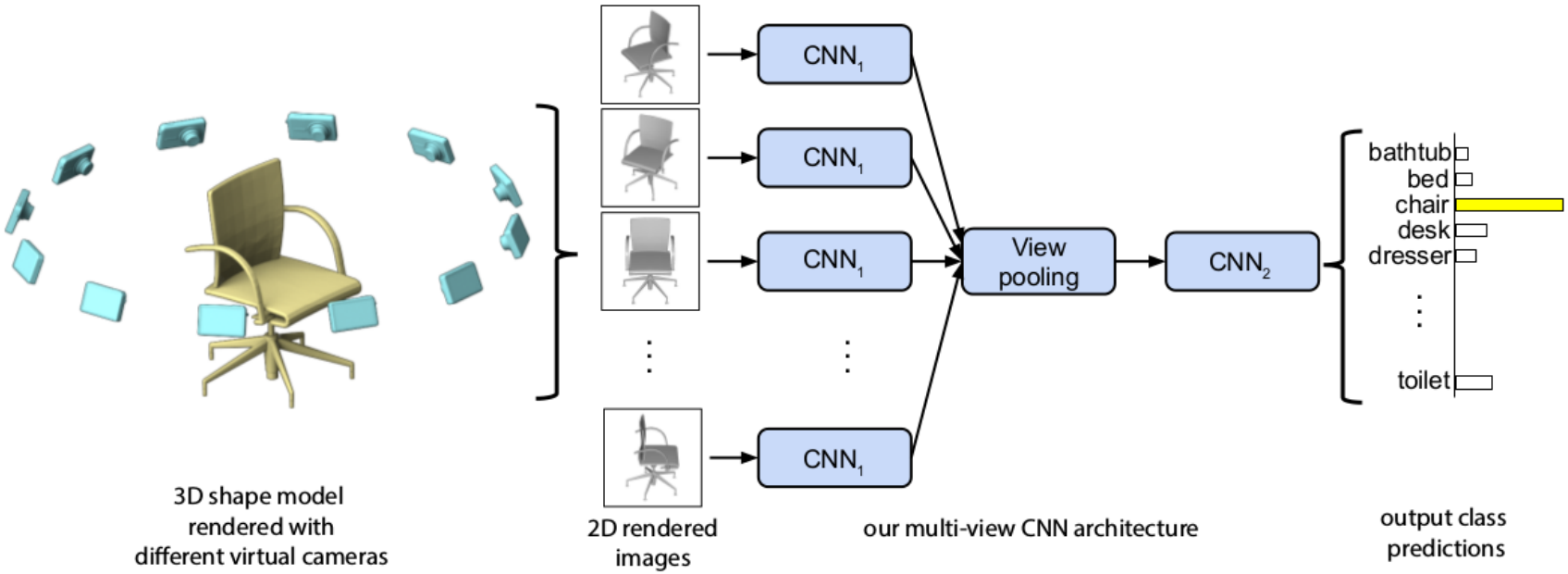
マルチビューベース



Multi-viewベースの3D物体認識（1/3）

Multi-view Convolutional Neural Networks for 3D Shape Recognition

H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. *IEEE ICCV,* 2015.



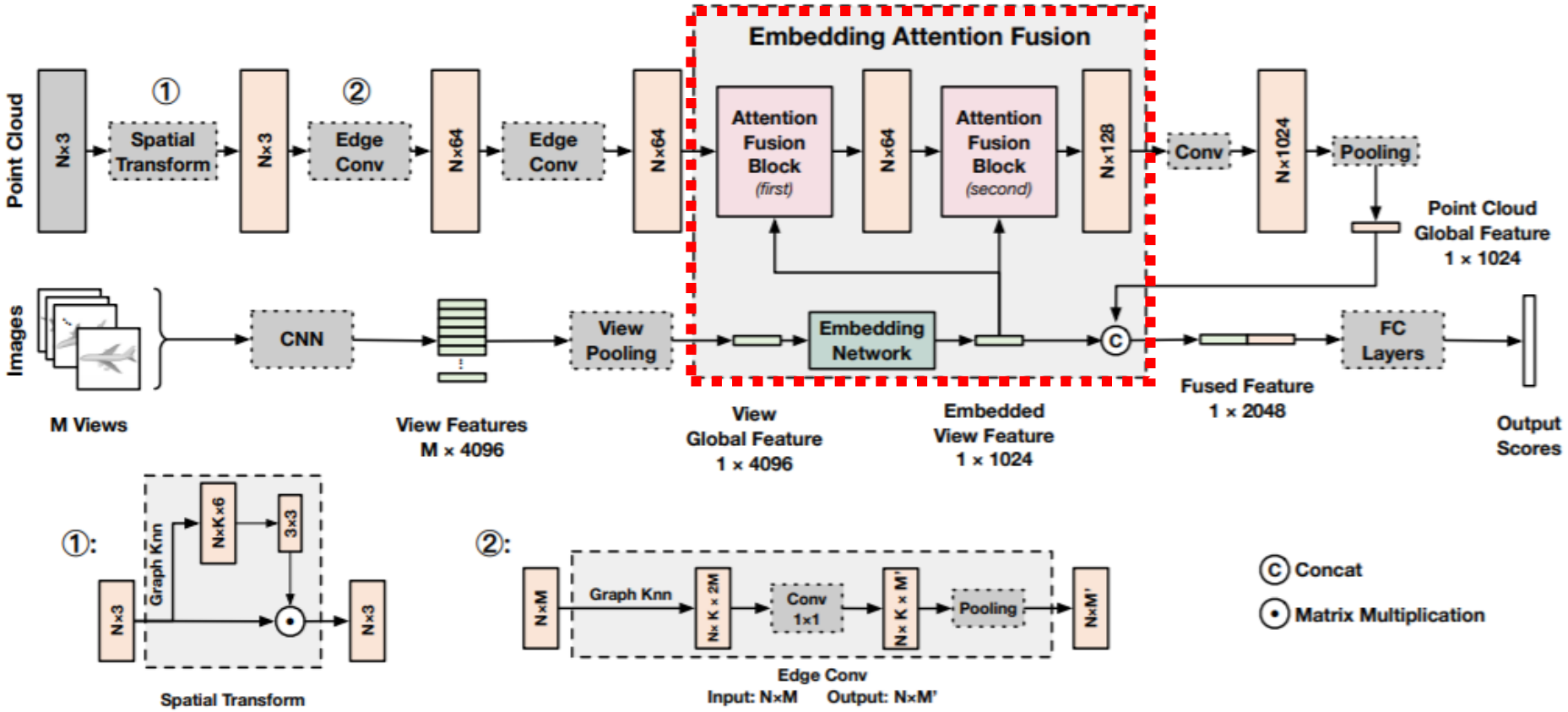
- VGG-Mアーキテクチャの CNN
- 中間層（Conv5）の後にView pooling層を入れて情報統合
- ModelNet40にて、ボクセルベースのShapeNetsと比べて8%性能向上 (77% → 85%)

Multi-viewベースの3D物体認識（2/3）

PVNet: A Joint Convolutional Network of Point Cloud and Multi-View for 3D Shape Recognition

Haoxuan You, Yifan Feng, Rongrong Ji, Yue Gao. ACM Multimedia, 2018.

マルチビュー画像と点群からAttention Fusionして精度改善

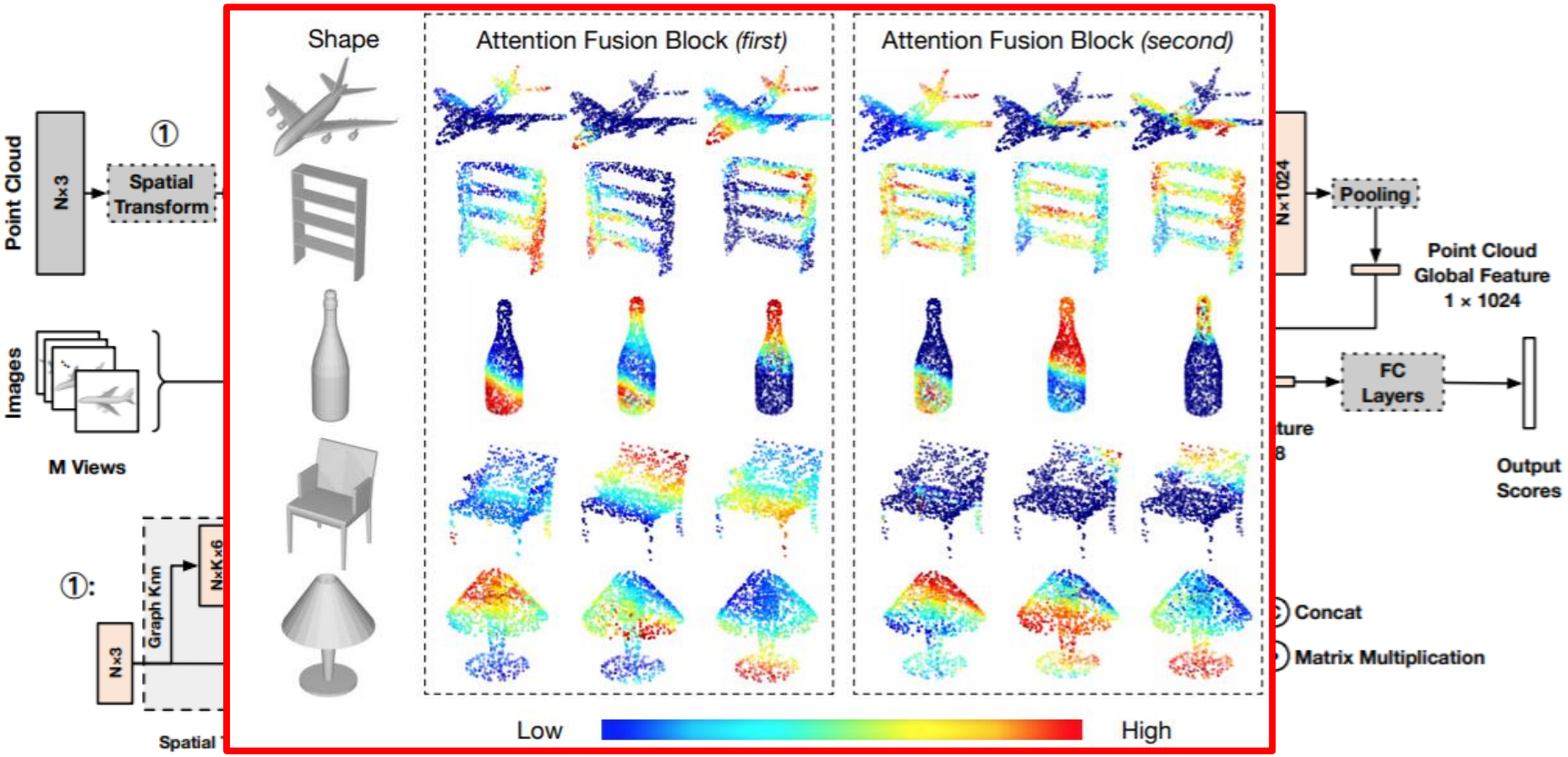


Multi-viewベースの3D物体認識（2/3）

PVNet: A Joint Convolutional Network of Point Cloud and Multi-View for 3D Shape Recognition

Haoxuan You, Yifan Feng, Rongrong Ji, Yue Gao. ACM Multimedia, 2018.

マルチビュー画像と点群からAttention Fusionして精度改善



ModelNet

http://modelnet.cs.princeton.edu/

- 40種類のModelNet40と
- 10種類のModelNet10がある。
- 2018/11/20現在

1位：RotationNet

Multi-viewベース

2位：PANORAMA-ENN

パノラマベース

3位：VRN Ensemble

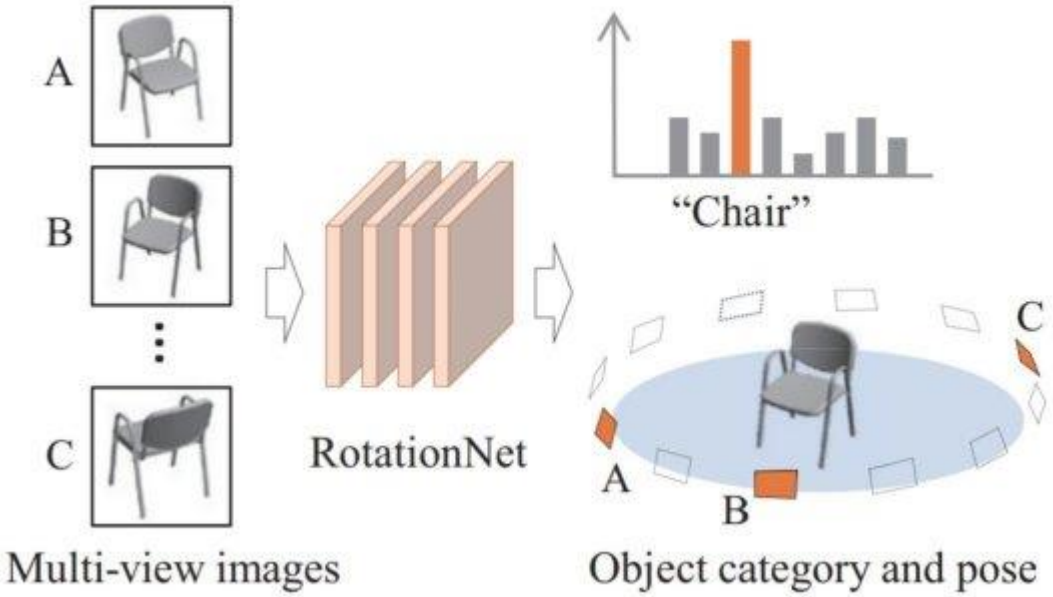
ボクセルベース ※精度は怪しい

Algorithm	ModelNet40 Classification (Accuracy)	ModelNet40 Retrieval (mAP)	ModelNet10 Classification (Accuracy)	ModelNet10 Retrieval (mAP)
PVNet[47]	93.2%	89.5%	-	-
GVCNN[46]	93.1%	85.7%	-	-
MLH-MV[45]	93.11%		94.80%	
MVCNN-New[44]	95.0%			
SeqViews2SeqLabels[43]	93.40%	89.09%	94.82%	91.43%
G3DNet[42]	91.13%		93.1%	
VSL [41]	84.5%		91.0%	
3D-CapsNets[40]	82.73%	70.1%	93.08%	88.44%
KCNet[39]	91.0%		94.4%	
FoldingNet[38]	88.4%		94.4%	
binVoxNetPlus[37]	85.47%		92.32%	
DeepSets[36]	90.3%			
3D-DescriptorNet[35]			92.4%	
SO-Net[34]	93.4%		95.7%	
Minto et al.[33]	89.3%		93.6%	
RotationNet[32]	97.37%		98.46%	
LonchaNet[31]			94.37	
Achlioptas et al. [30]	84.5%		95.4%	
PANORAMA-ENN [29]	95.56%	86.34%	96.85%	93.28%
3D-A-Nets [28]	90.5%	80.1%		
Soltani et al. [27]	82.10%			
Arvind et al. [26]	86.50%			
LonchaNet [25]			94.37%	
3DmFV-Net [24]	91.6%		95.2%	
Zanuttigh and Minto [23]	87.8%		91.5%	
Wang et al. [22]	93.8%			
ECC [21]	83.2%		90.0%	
PANORAMA-NN [20]	90.7%	83.5%	91.1%	87.4%
MVCNN-MultiRes [19]	91.4%			
FPNN [18]	88.4%			
PointNet[17]	89.2%			
Klokov and Lempitsky[16]	91.8%		94.0%	
LightNet[15]	88.93%		93.94%	
Xu and Todorovic[14]	81.26%		88.00%	
Geometry Image [13]	83.9%	51.3%	88.4%	74.9%
Set-convolution [11]	90%			
PointNet [12]			77.6%	
3D-GAN [10]	83.3%		91.0%	
VRN Ensemble [9]	95.54%		97.14%	
ORION [8]			93.8%	
FusionNet [7]	90.8%		93.11%	
Pairwise [6]	90.7%		92.8%	
MVCNN [3]	90.1%	79.5%		
GIFT [5]	83.10%	81.94%	92.35%	91.12%
VoxNet [2]	83%		92%	
DeepPano [4]	77.63%	76.81%	85.45%	84.18%
3DShapeNets [1]	77%	49.2%	83.5%	68.3%

Multi-viewベースの3D物体認識（3/3）

RotationNet: Joint Object Categorization and Pose Estimation Using Multiviews from Unsupervised Viewpoints

Asako Kanezaki, Yasuyuki Matsushita, and Yoshifumi Nishida. *IEEE CVPR*, 2018.

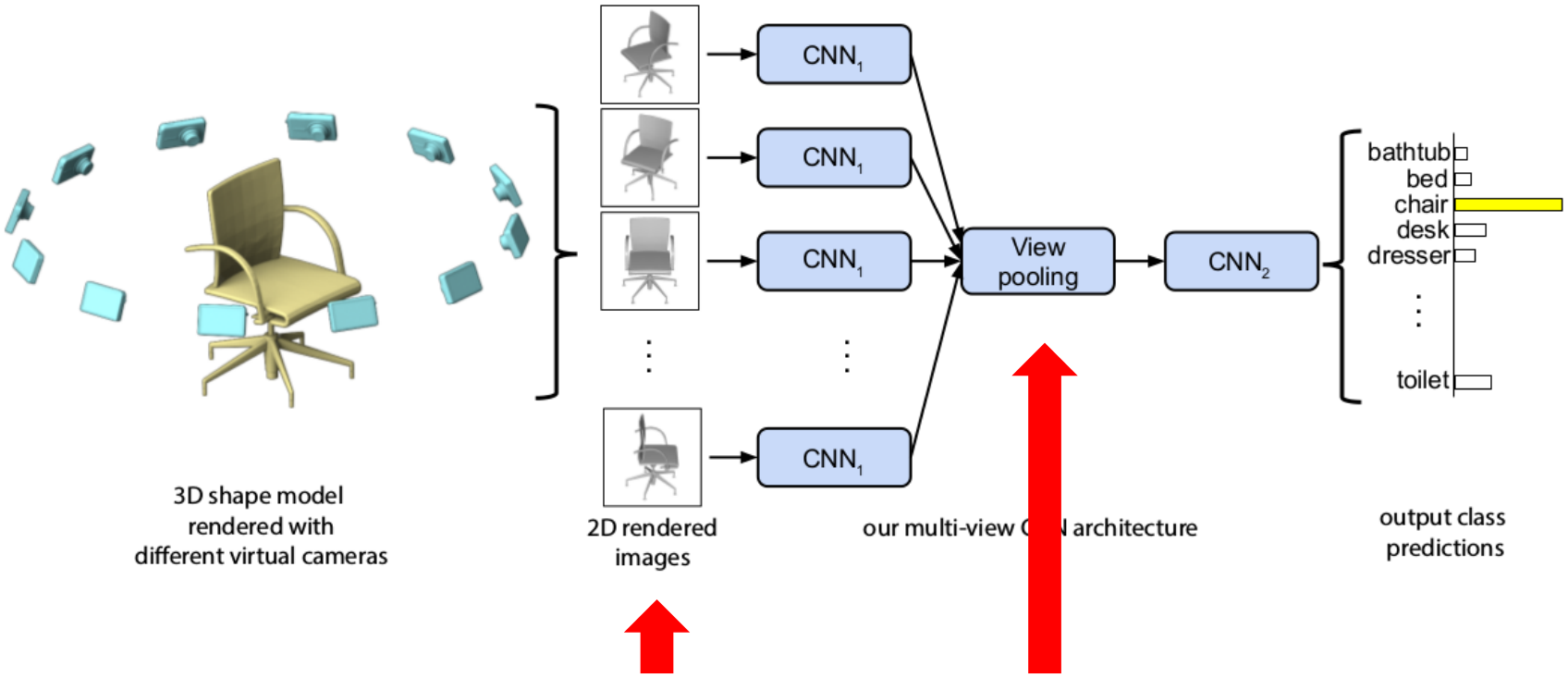


- 一連のマルチビュー画像を入力とするCNN。
- 物体のカテゴリと姿勢（各画像の対応する視点）を同時に推定する。
- 学習画像の視点情報は教示不要。（自動アラインメント機能）
- テスト時に入力するマルチビュー画像は1枚～数枚でOK。
- ModelNet10, 40でSOTA、SHREC'17のトラック1とトラック3で優勝。

RotationNet – 背景 –

Multi-view Convolutional Neural Networks for 3D Shape Recognition

H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. *IEEE ICCV,* 2015.

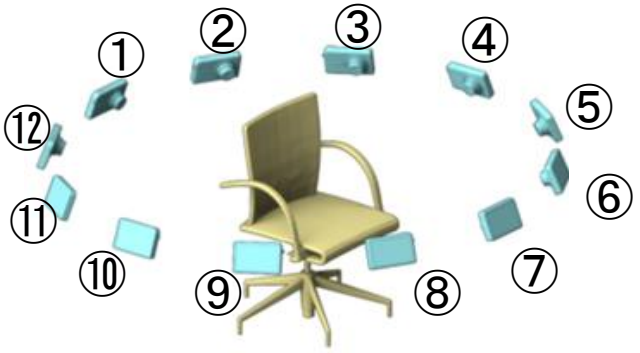


① テスト時も学習時と同じ数だけの
マルチビュー画像を同時入力せねばならない

② （回転不変性確保のため）
画像の順序情報を捨てている

RotationNet – モチベーションと課題 –

画像の順序を保持して、順序依存表現にすれば性能が上がる！



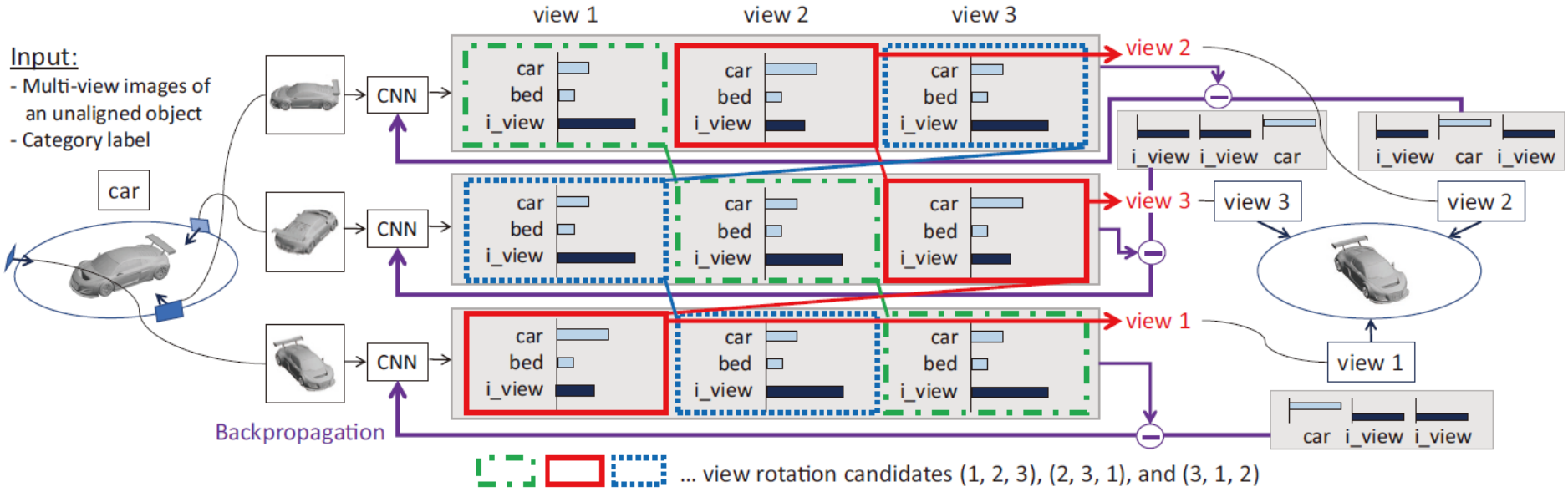
【課題1】各画像がどの視点に対応するかを推定せねばならない

⇒⑧？

【課題2】データベース内の物体の姿勢が揃っていない（ex. ModelNet）
自動的に向きを揃えなければならない

【課題3】テスト時に全ての画像が観測できない場合がある（ex. オクルージョン）
テスト時は1枚～任意枚数の入力画像で認識できなければならない

RotationNet – 提案手法 –



Forward:

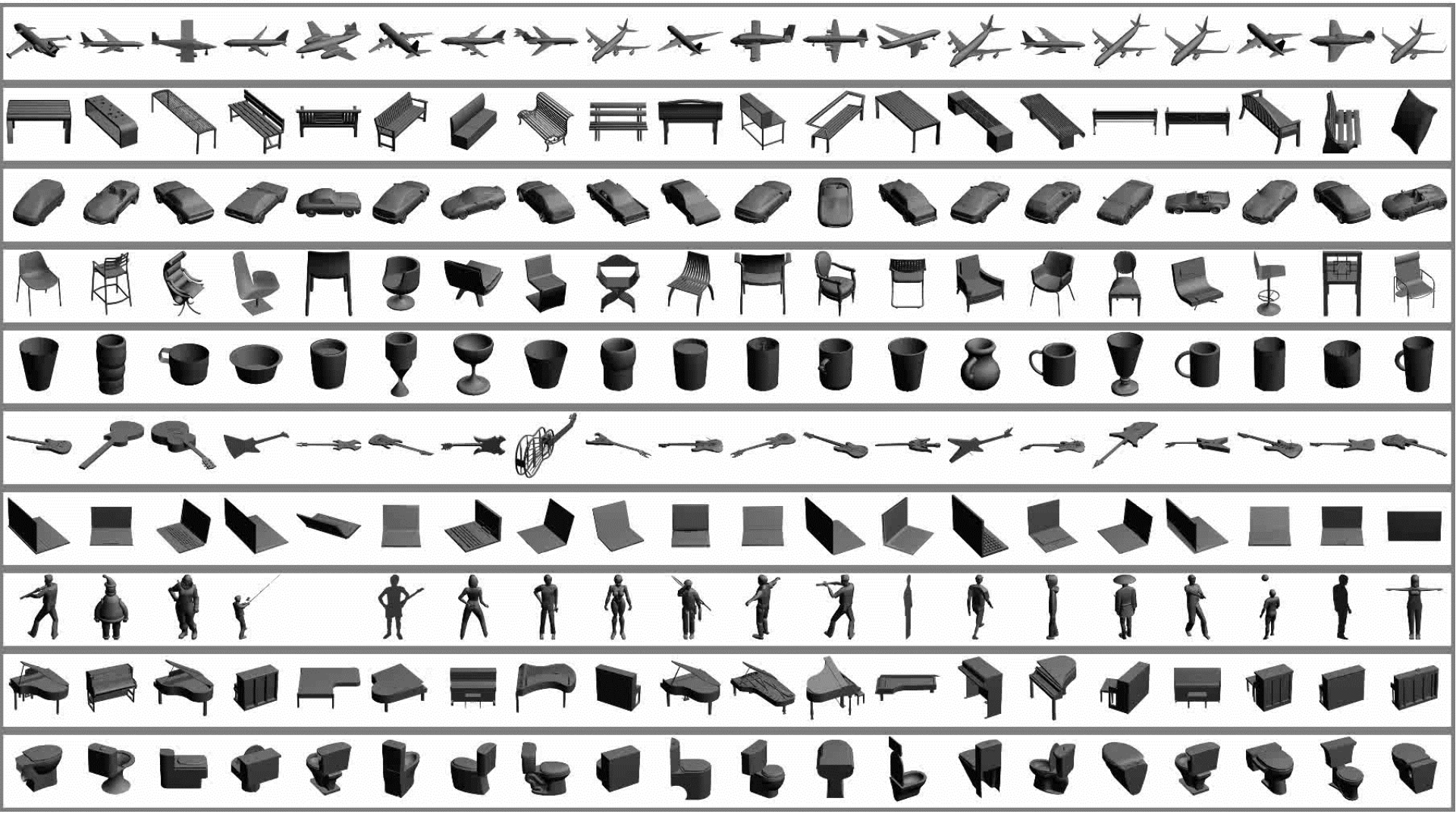
- 各画像に対して、各視点における物体カテゴリ尤度を出力する。
  ※物体カテゴリ尤度＝Nクラスのうちどれかあるいはどれでもない（別の視点から撮られた画像である； incorrect view）の$N+1$クラスの識別スコア
- 視点の個数をMとすると、M個の$M(N+1)$次元ベクトルを出力する。
- 掛け合わせたときの正解物体スコアが最大になるよう視点を割り当てる。

Backward:

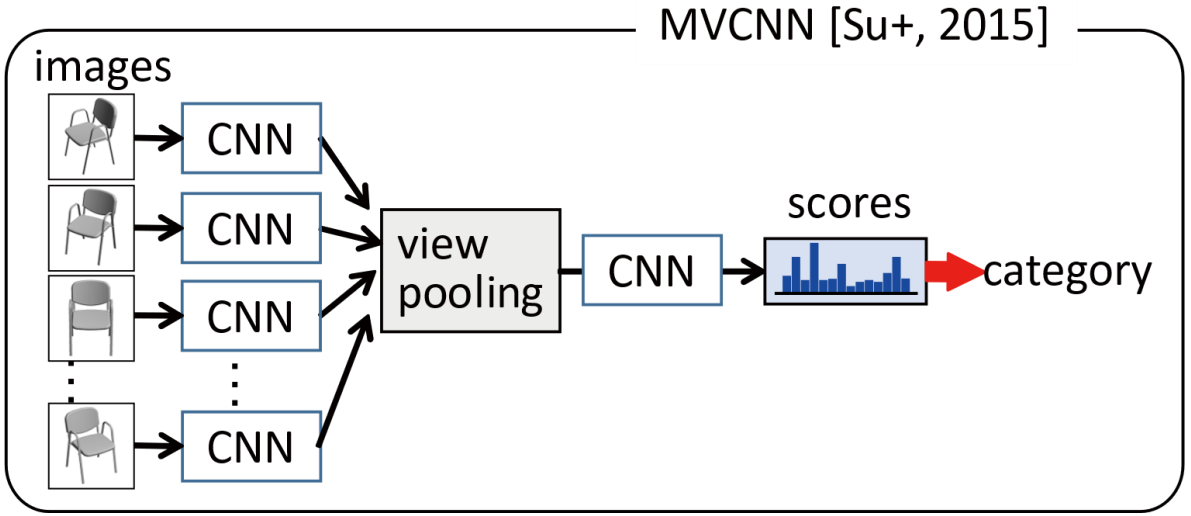
- 割り当てられた視点に対応する正解物体カテゴリ尤度が1になる勾配を求めてSGDする。

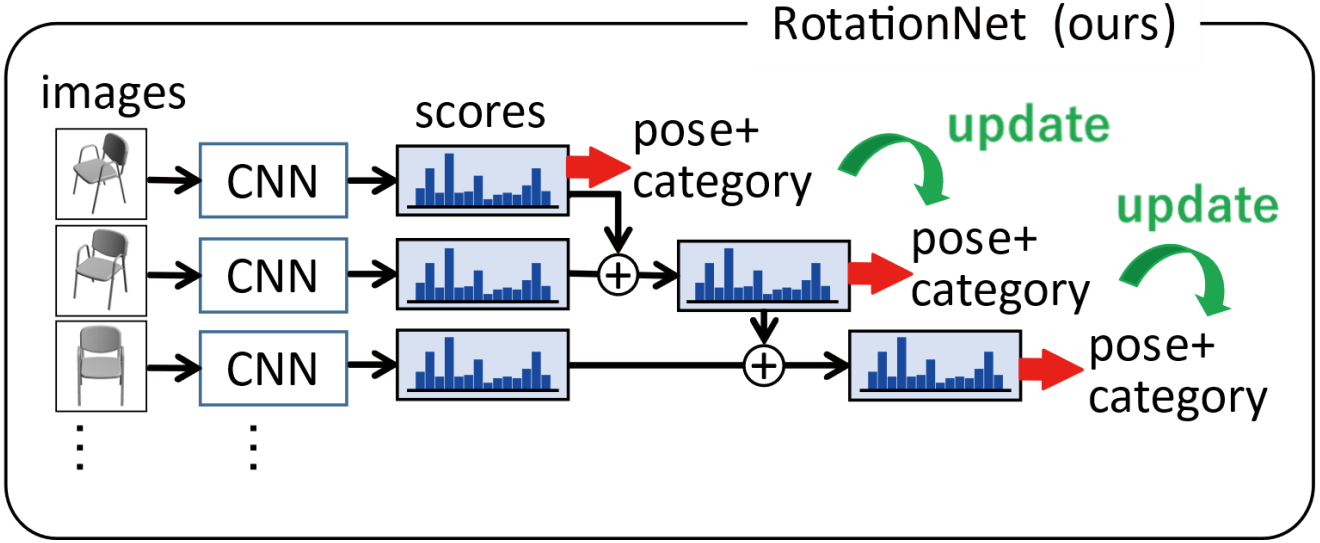
RotationNet – 学習時の自動姿勢調整の様子（動画）–

Iteration 0



RotationNet – 逐次的な画像入力が可能 –





RotationNet: Joint Object Categorization and Pose Estimation Using Multiviews from Unsupervised Viewpoints

CVPR 2018

Asako Kanezaki, Yasuyuki Matsushita, and Yoshifumi Nishida

SHREC2017 - 3D Shape Retrieval Contest 2017

Eurographics 2017 Workshop on 3D Object Retrieval, http://liris.cnrs.fr/eg3dor2017/

- **RGBD物体データからCADモデルを検索**
- 3Dハンドジェスチャー認識
- **大規模3D形状検索**
- タンパク質形状識別
- 非剛体玩具の点群形状検索
- 欠陥のある非剛体形状検索
- レリーフパターン検索

7トラック中2トラックに
参加し一位を獲得！

両トラックでRotationNetを使用

ポイント:

検索タスクだけど物体識別が使えた！

- カテゴリラベル付きのTrain, Valデータが配られた。
- テストデータのカテゴリを識別して、クエリの（推定）カテゴリに対して識別スコアの高い順に物体を提示するという戦法をとった。

トラック1：RGBD物体データからCADモデルを検索

クエリのRGBデータに対し検索結果に同じカテゴリの物体がどれだけ含まれるかを競う

- CADモデルデータを学習した識別器を、RGBDデータでFine-tuningすることで性能が向上した。
- 優勝！

Method	Precision	Recall	F1	mAP	NDCG	Tier-1	Tier-2
Kanezaki-Single	0.792	0.792	0.792	0.792	0.792	0.792	0.792
Kanezaki-Thresh	0.793	0.799	0.794	0.794	0.796	0.794	0.794
Kanezaki-1000	**0.820**	**0.820**	**0.820**	**0.833**	**0.805**	**0.824**	**0.824**
Tang-3DCNN	0.769	0.769	0.769	0.749	0.774	0.769	0.769
Tang-MVCNN	0.727	0.727	0.727	0.710	0.735	0.727	0.727
Tang-Fuse	0.759	0.759	0.759	0.746	0.763	0.759	0.759
Tang-CDTNN	0.672	0.672	0.672	0.649	0.714	0.672	0.672
Truong-2D	0.740	0.740	0.740	0.740	0.740	0.740	0.740
Truong-3D	0.487	0.487	0.487	0.487	0.487	0.487	0.487
Tran-1	0.703	0.703	0.703	0.703	0.703	0.703	0.703
Tran-2	0.690	0.690	0.690	0.676	0.695	0.690	0.690
Tran-3	0.691	0.691	0.691	0.691	0.691	0.691	0.691
Tran-4	0.689	0.689	0.689	0.675	0.692	0.689	0.689
Li	0.105	0.320	0.145	0.062	0.476	0.120	0.100
Tashiro	0.141	0.472	0.198	0.149	0.552	0.188	0.144

↑ 学習有

学習無 ↓

トラック3：大規模3D形状検索（1/2）

クエリのCADモデルに対し検索結果に同じカテゴリの物体がどれだけ含まれるかを競う

- タスク1：姿勢が揃っている、タスク2：姿勢がバラバラ
- タスク1の方で優勝！

Dataset	Method	microALL P@N	R@N	F1@N	mAP	NDCG	macroALL P@N	R@N	F1@N	mAP	NDCG
test_normal	Kanezaki_RotationNet	0.810	0.801	0.798	0.772	0.865	0.602	0.639	0.590	0.583	0.656
	Zhou_Improved_GIFT	0.786	0.773	0.767	0.722	0.827	0.592	0.654	0.581	0.575	0.657
	Tatsuma_ReVGG	0.765	0.803	0.772	0.749	0.828	0.518	0.601	0.519	0.496	0.559
	Furuya_DLAN	0.818	0.689	0.712	0.663	0.762	0.618	0.533	0.505	0.477	0.563
	Thermos_MVFusionNet	0.743	0.677	0.692	0.622	0.732	0.523	0.494	0.484	0.418	0.502
	Deng_CM-VGG5-6DB	0.418	0.717	0.479	0.540	0.654	0.122	0.667	0.166	0.339	0.404
	Li_ZFDR	0.535	0.256	0.282	0.199	0.330	0.219	0.409	0.197	0.255	0.377
	Mk_DeepVoxNet	0.793	0.211	0.253	0.192	0.277	0.598	0.283	0.258	0.232	0.337
	SHREC16-Bai_GIFT	0.706	0.695	0.689	0.640	0.765	0.444	0.531	0.454	0.447	0.548
	SHREC16-Su_MVCNN	0.770	0.770	0.764	0.735	0.815	0.571	0.625	0.575	0.566	0.640

トラック3：大規模3D形状検索（2/2）

クエリのCADモデルに対し検索結果に同じカテゴリの物体がどれだけ含まれるかを競う

- タスク2はPoint Cloudベースが優勝（やはり回転に強い…）
- ただしView数を増やすとRotationNetが勝つ（コンペ後の追加実験の結果。）

Dataset	Method	microALL P@N	microALL R@N	microALL F1@N	microALL mAP	microALL NDCG	macroALL P@N	macroALL R@N	macroALL F1@N	macroALL mAP	macroALL NDCG
test_perturbed	Furuya_DLAN	0.814	0.683	0.706	0.656	0.754	0.607	0.539	0.503	0.476	0.560
	Tatsuma_ReVGG	0.705	0.769	0.719	0.696	0.783	0.424	0.563	0.434	0.418	0.479
	Zhou_Improved_GIFT	0.660	0.650	0.643	0.567	0.701	0.443	0.508	0.437	0.406	0.513
	Kanezaki_RotationNet	0.655	0.652	0.636	0.606	0.702	0.372	0.393	0.333	0.327	0.407
	Deng_CM-VGG5-6DB	0.412	0.706	0.472	0.524	0.642	0.120	0.659	0.164	0.329	0.395
	Li_ZFDR	0.496	0.234	0.258	0.172	0.303	0.199	0.373	0.179	0.215	0.336
	Mk_DeepVoxNet	0.690	0.012	0.020	0.009	0.043	0.546	0.052	0.052	0.047	0.109
	SHREC16-Bai_GIFT	0.678	0.667	0.661	0.607	0.735	0.414	0.496	0.423	0.412	0.518
	SHREC16-Su_MVCNN	0.632	0.613	0.612	0.535	0.653	0.405	0.484	0.416	0.367	0.459

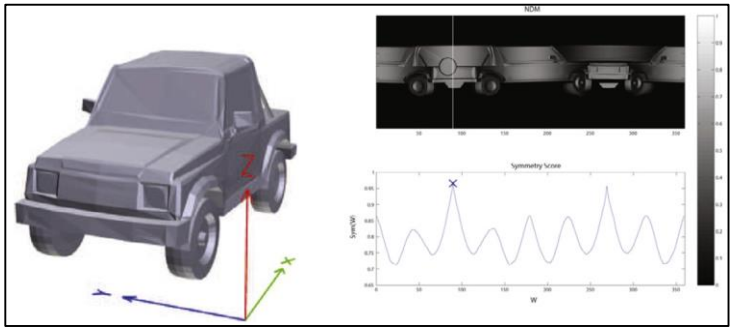
Multi-viewベースの3D物体認識（まとめ）

- 現状、性能が一番高い。
- 実装が簡単（複数画像をバッチに押し込むだけ）。
- 連続画像（動画像）にも適用可能。
- 見たことない視点からの認識に弱い。
  （視点の数を増やすと性能が上がる。）

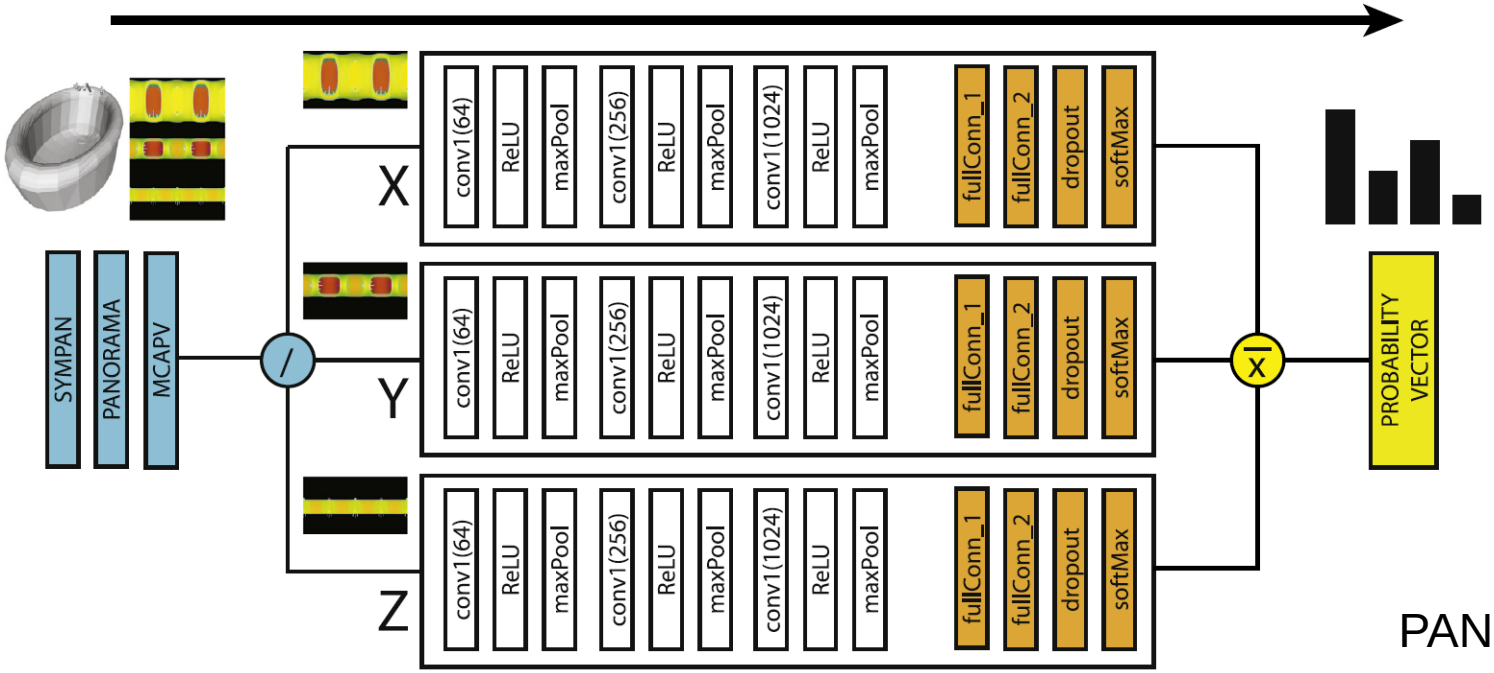
【その他の方法1】

Ensemble of PANORAMA-based Convolutional Neural Networks for 3D Model Classification and Retrieval

K. Sfikas, I. Pratikakis and T. Theoharis. *Computers and Graphics*, 2018.



- 主成分分析でx, y, z軸を決定する。
- z軸方向を縦として物体を囲む円柱を立てる。
- 円柱に物体表面上の点を投影する。
- 左右対称性が最大の点を基準とする。
- x, y, z各軸に対して勾配等3チャンネル画像を作成。
- CNNに入力・スコアをlate fusionする。

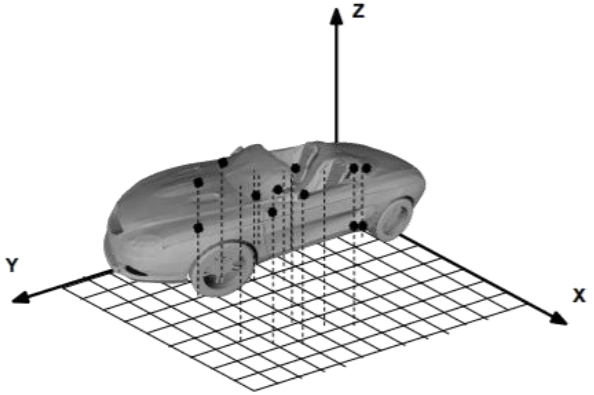


PANORAMA-ENN

【その他の方法2】

Learning 3D Shapes as Multi-Layered Height-maps using 2D Convolutional Networks

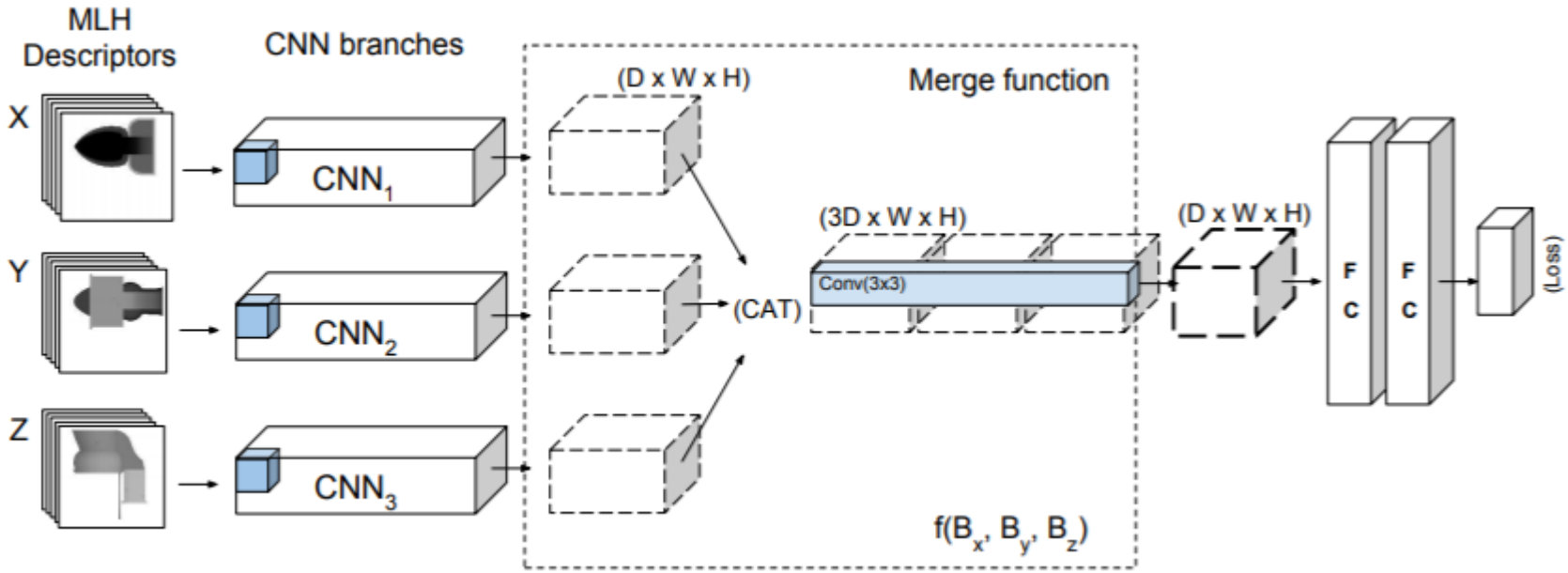
Sarkar, K., Hampiholi, B., Varanasi, K. and Stricker, D. *ECCV*, 2018.



k段階の高さを抽出したNxNxk次元のMLH記述子を提案

$$M_{[p,q,i]} \leftarrow ((i-1)/(k-1) * 100)^{th} \text{ percentile of } P_{pq}$$

P_{pq}: 点の高さのセット



3D物体認識の分類（本講演）

RGBDベース

- MMSS [ICCV'15]
- HHA [ECCV'14]
- Depth CNNs for RGB-D scene recognition [AAAI'17]
- Augmented Autoencoder [ECCV'18]

Point Cloudベース

- PointNet [CVPR'17]
- SO-Net [CVPR'18]
- Attentional ShapeContextNet [CVPR'18]
- Tangent Convolutions [CVPR'18]
- PPFNet [CVPR'18]

Voxelベース

- 3D ShapeNets [CVPR'15]
- ORION [BMVC'17]
- PointGrid [CVPR'18]
- CubeNet [ECCV'18]

Multi-viewベース

- MVCNN [ICCV'15]
- PVNet [ACM MM'18]
- RotationNet [CVPR'18]

その他のアプローチ：PANORAMA-ENN, MLH

3D物体認識の分類（所感）

RGBDベース

- 2.5次元。
- Depth画像はHHAコーディング。
- 姿勢推定によく使われる。

Point Cloudベース

- 回転不変な局所特徴抽出が肝。
- 大域特徴への統合も肝。
- 物体の回転に強い。
- 性能は高くない。
- パーツセグメンテーションに応用できる。

Voxelベース

- 低解像度のため性能が高くない。
- 回転に弱い。

Multi-viewベース

- 性能が高い。
- 実装が簡単。
- 連続画像（動画像）にも適用可能。
- 見たことない視点からの認識に弱い。

その他のアプローチ：パノラマベースも認識精度が高い。ただ、姿勢は揃える必要あり。