Technische Universität München

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Morphing and Sampling Network for Dense Point Cloud Completion (MSN)

Minghua Liu, Lu Sheng, Sheng Yang, Jing Shao, Shi-Min Hu
Presented at the AAAI Conference on Artificial Intelligence, 2020, New York

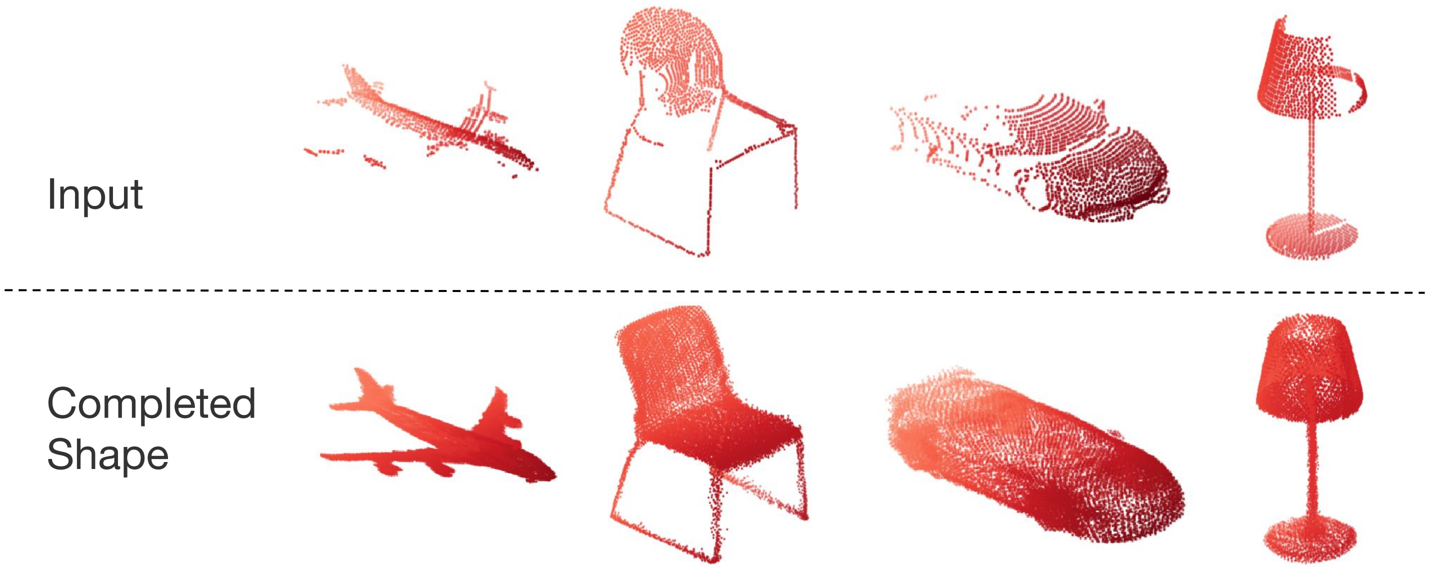
Student: Lukas Legner
Tutor: Yida Wang

Technische Universität München
Johns Hopkins Whiting School of Engineering

Presentation Structure

Motivation

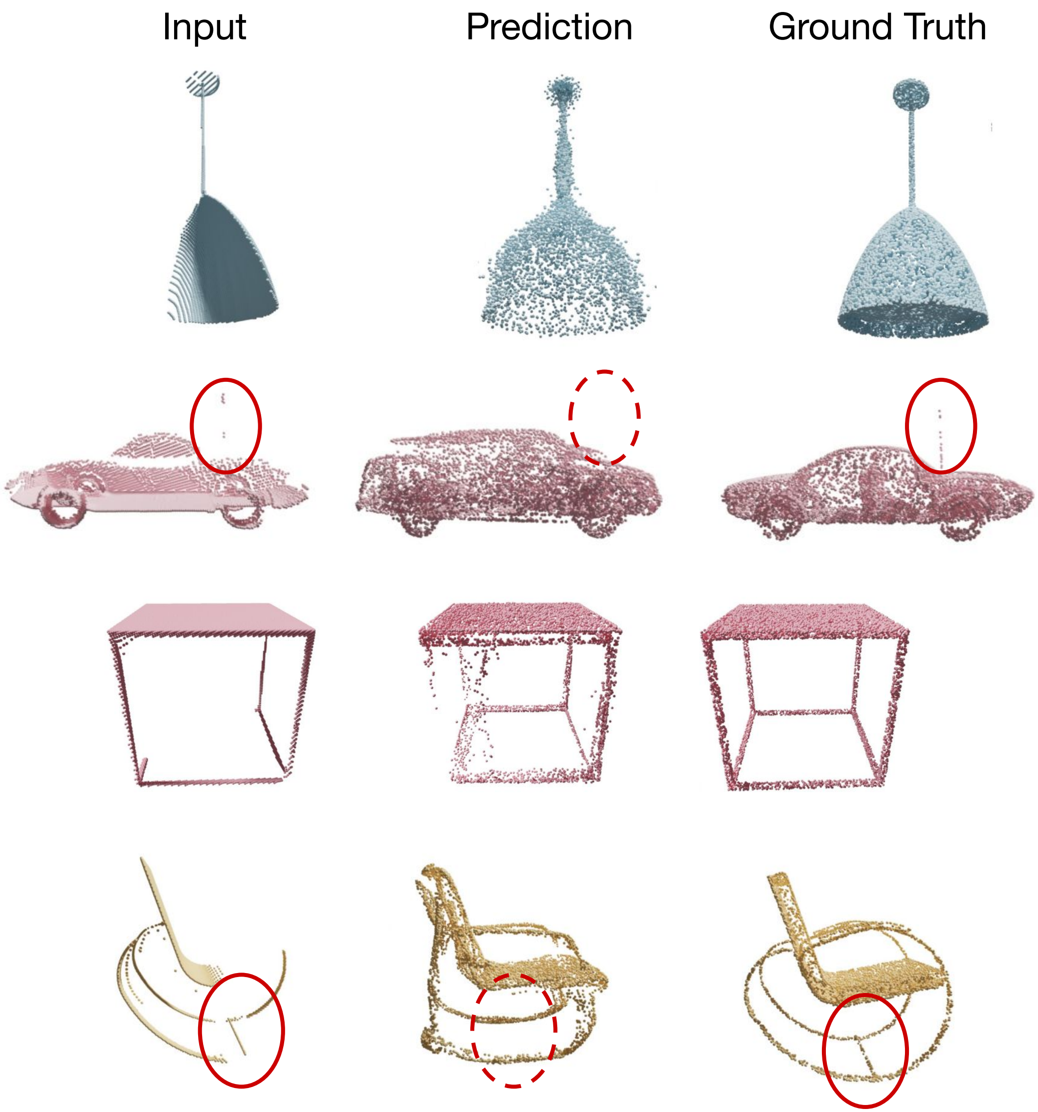
- Real-world scan data is imperfect:
  - Limited view angles
  - Limited scanner capacity
- Point cloud completion to improve performance of subsequent tasks like
  - Shape classification
  - Point cloud registration
  - Semantic segmentation



Point cloud completion on ShapeNet dataset, adapted from [2]

Challenges in Point Cloud Completion

1. Continuous and smooth surfaces
2. Capture fine details of the object
3. Locally even distribution of points
4. Preserve existing structures from the input

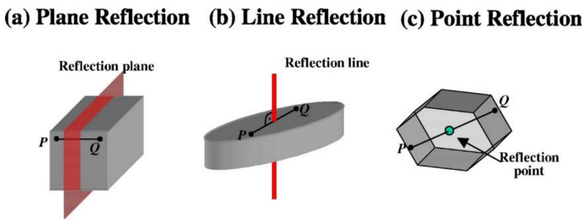


Images adapted from [1]

Related Work - 3D Shape Completion

How to complete a 3D shape?
Common approaches:

- Geometry-based
  - Interpolation
  - Assumptions about geometry (for example symmetry)

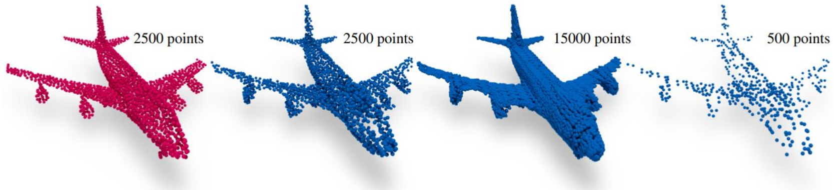


Reflection symmetries [3]

- Example-based
  - Partial shape database: Deform+assemble

Model shape database [4]

- Learning-based
  - Voxel grids + 3D convs
  - Polygon meshes + graph-based convs
  - **Point clouds**

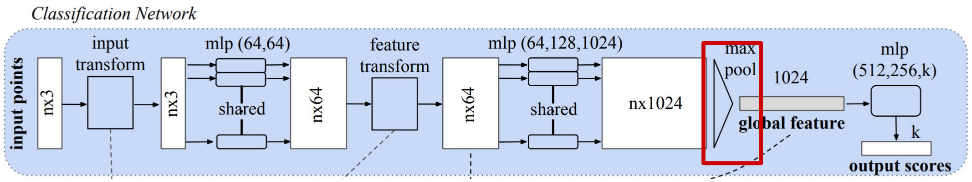


Point clouds of varying densities [5]

Related Work - Point Cloud Based

- PointNet (2017) [6]
  - Directly consumes point cloud
  - Permutation invariance
  - Used as backbone for point cloud encoding

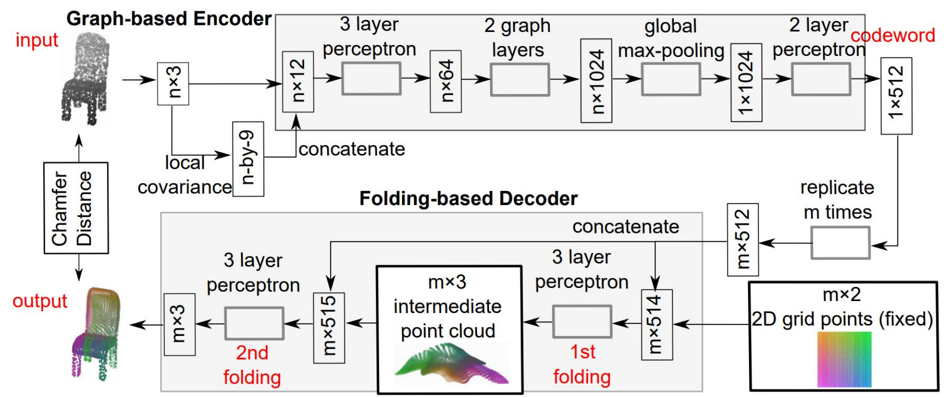
PointNet encoder structure:
Input → 1DConvs + feature transforms → Max pool → Feature Vector



PointNet architecture, classifier. From [6]

- FoldingNet (2018) [7]
  - Encoder: Encode input point cloud into a feature vector
  - Decoder: Create 2D grids of points and deform them into the original object

FoldingNet network structure:
Input → Encoder → Feature Vector ('codeword') → Decoder -> Output



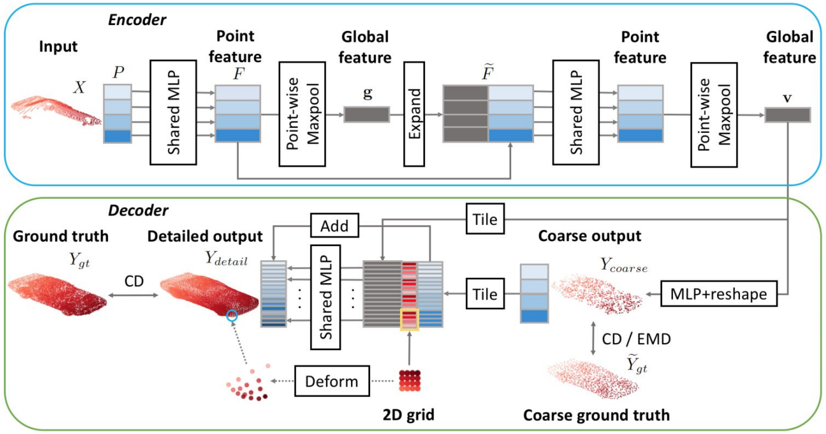
FoldingNet architecture, from [7]

Idea:
Deform 2D grid into 3D object surface
✓ Smooth surfaces

Related Work - Point Cloud Based

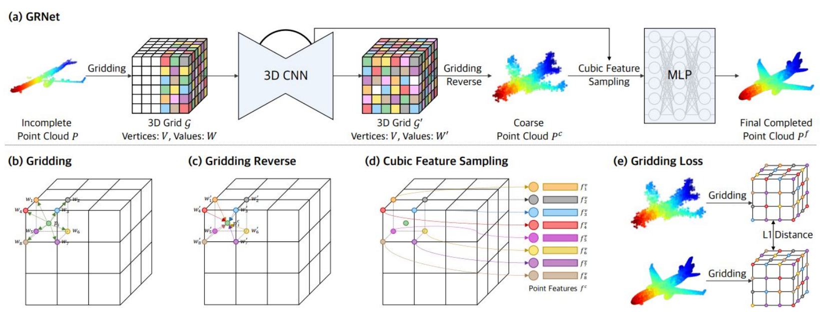
- Point Completion Network (2018) [8]
  - Encoder: similar to FoldingNet
  - Decoder: Create coarse point cloud, then refine it into detailed point cloud

Idea:
Generate coarse point cloud then refine into detailed shape
✓ Fine details



PCN architecture, from [8]

- GRNet (2020) [9]
  - Voxel-based
  - *Gridding* and *Gridding-Reverse* layers to convert between point cloud and 3D grid



GRNet architecture, from [9]

Morphing and Sampling Network (MSN)

What we have so far

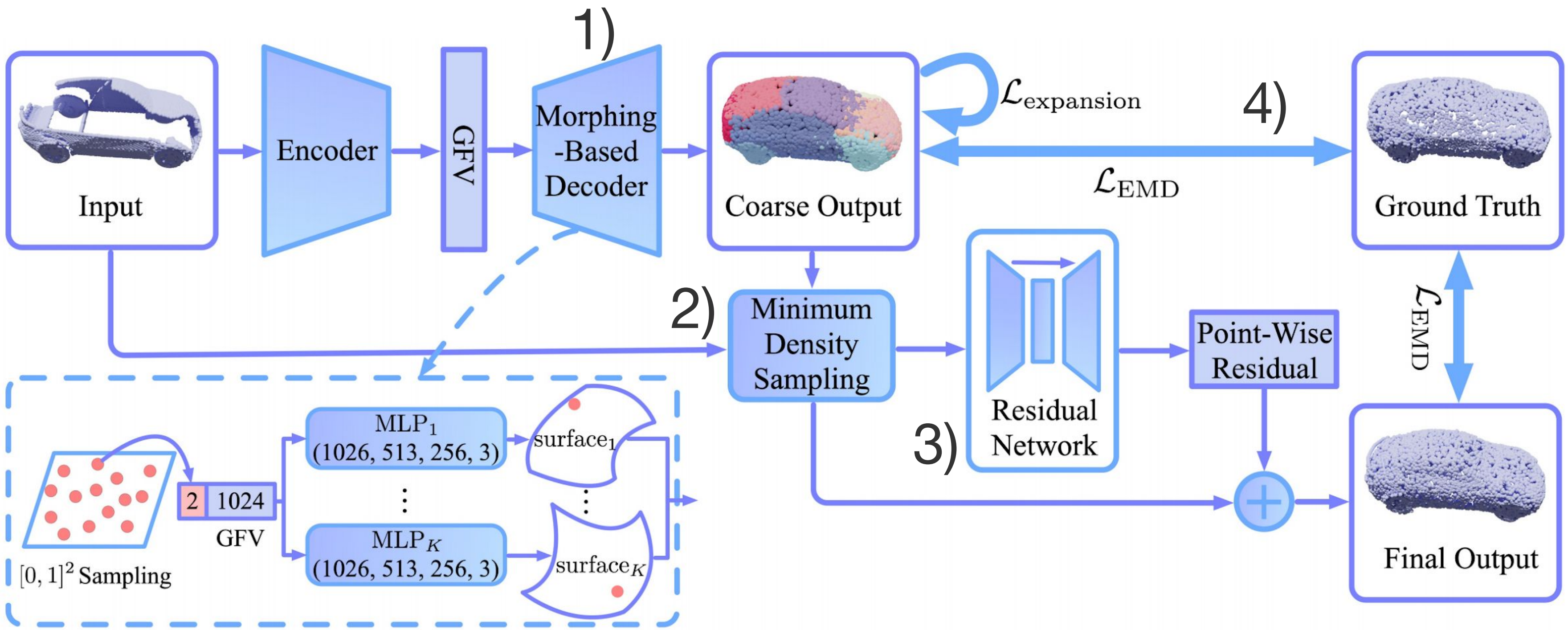
1. Continuous and smooth surfaces ✓ Morphing-Based Decoder
2. Capture fine details of the object ✓ Refinement
3. Locally even distribution of points
4. Preserve existing structures from the input

Contributions of the paper

- Novel approach for point cloud estimation
- Addition of an expansion penalty for surface elements
- Novel sampling algorithm for point clouds with evenly distributed results
- Implementation of an Earth mover's distance (EMD) approximation

The Network Architecture in 4 Parts

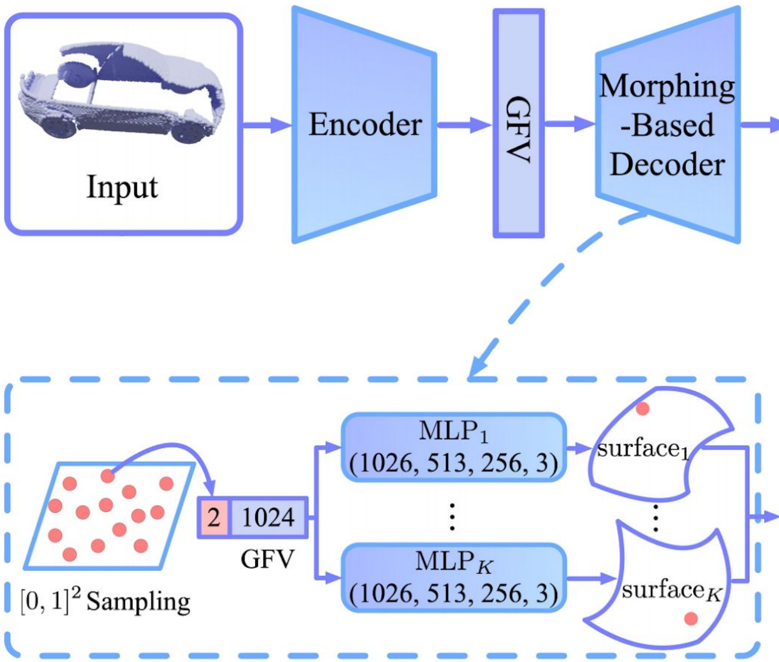
Overall Architecture



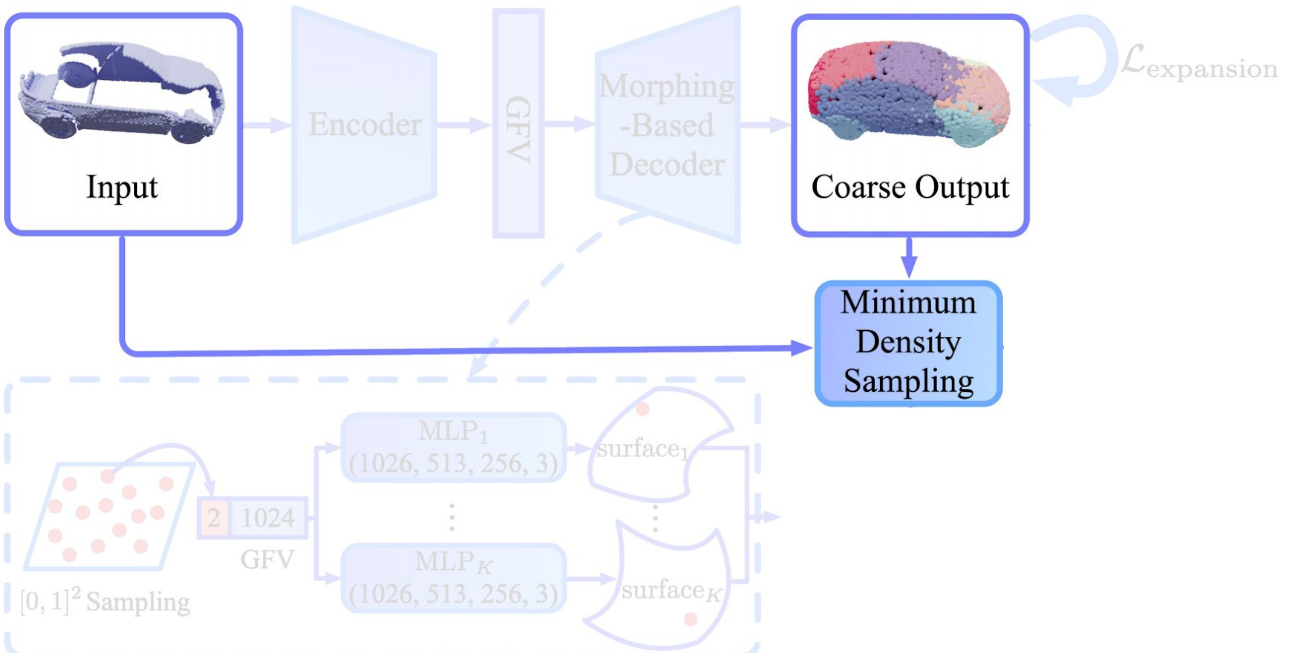
MSN architecture, from [1]

Structure of the network:

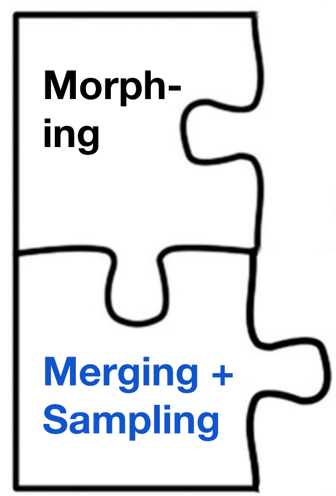
Input → 1) Morphing → Coarse Output→ 2) Merging+Sampling
→ 3) Refinement → Final Output → 4) Loss


Input
Encoder
GFV
Morphing
-Based
Decoder
Coarse Output
MLP$_1$
(1026, 513, 256, 3)
MLP$_K$
(1026, 513, 256, 3)
surface$_1$
surface$_K$
2
1024
GFV
$[0, 1]^2$ Sampling

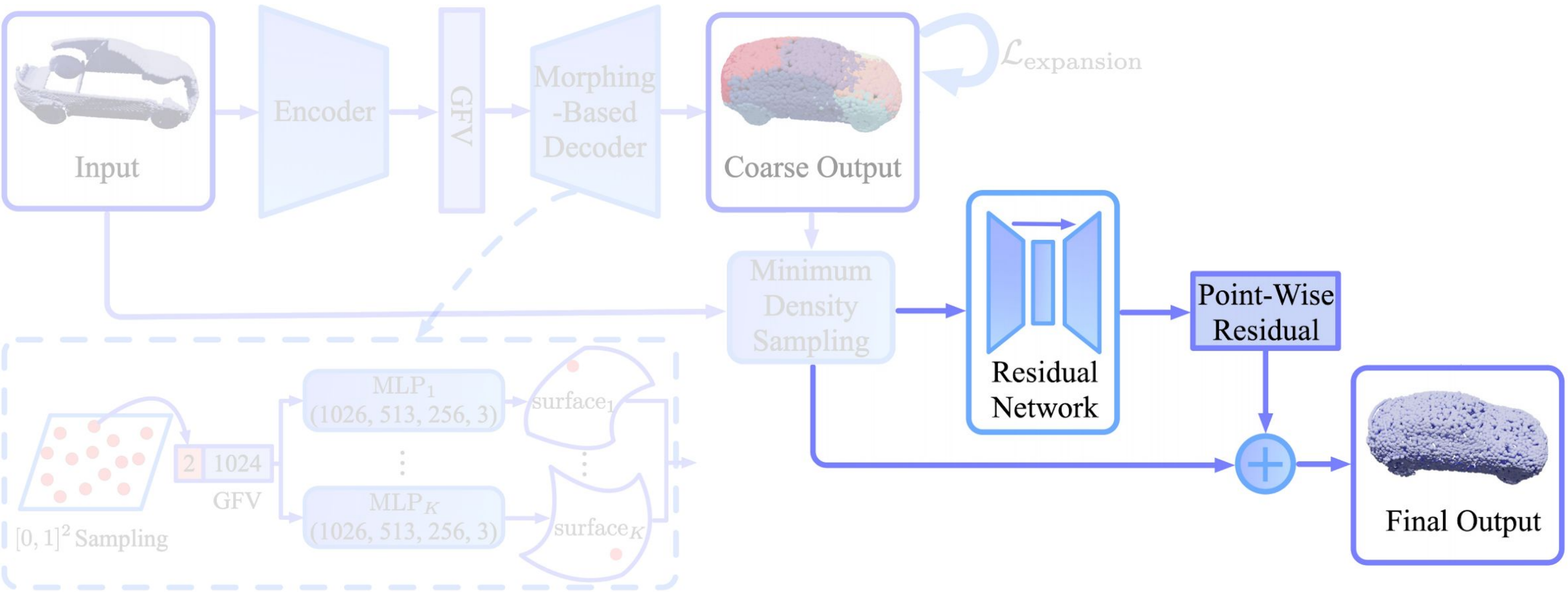

Morph-
ing



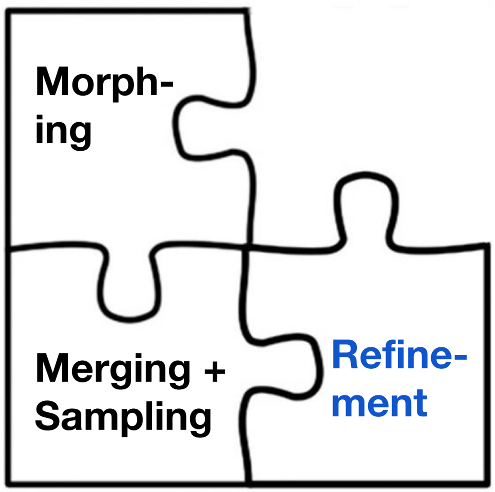
Input
Encoder
GFV
Morphing
-Based
Decoder
Coarse Output
$\mathcal{L}_{expansion}$
Minimum
Density
Sampling
MLP$_1$
(1026, 513, 256, 3)
surface$_1$
2
1024
GFV
MLP$_K$
(1026, 513, 256, 3)
surface$_K$
$[0, 1]^2$ Sampling



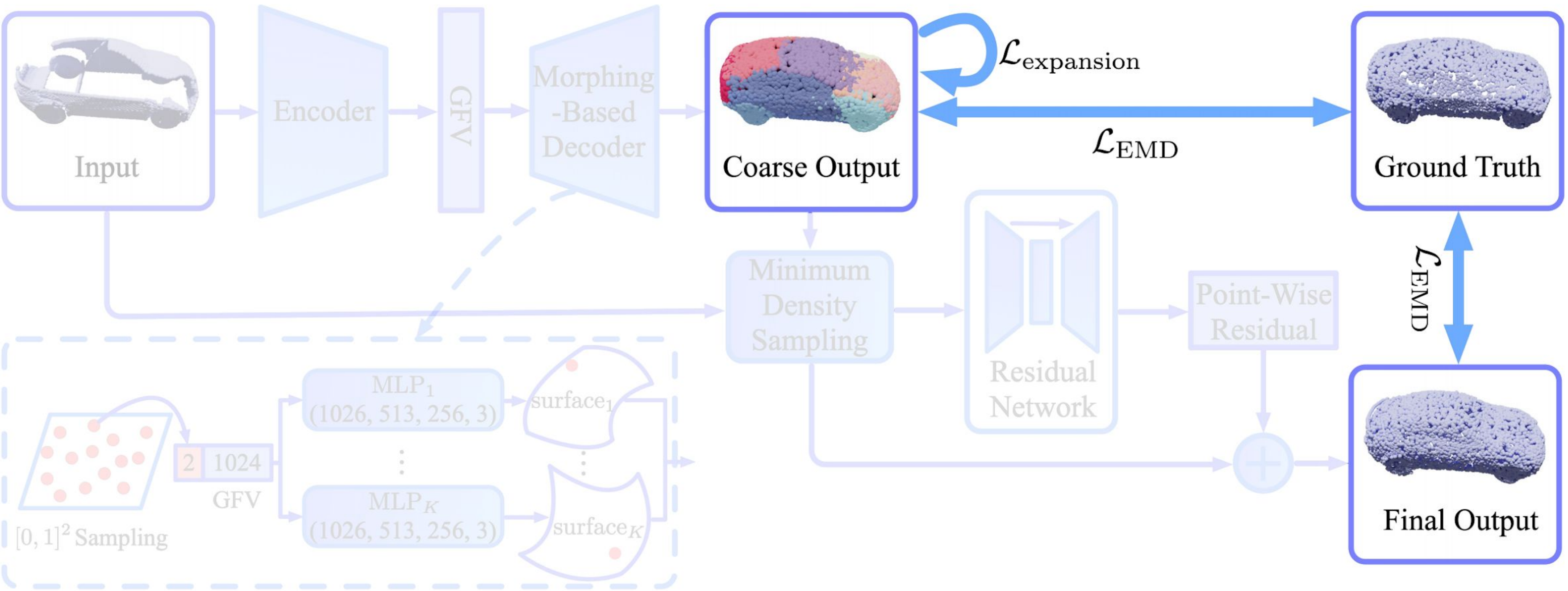
Morph-
ing
Merging +
Sampling

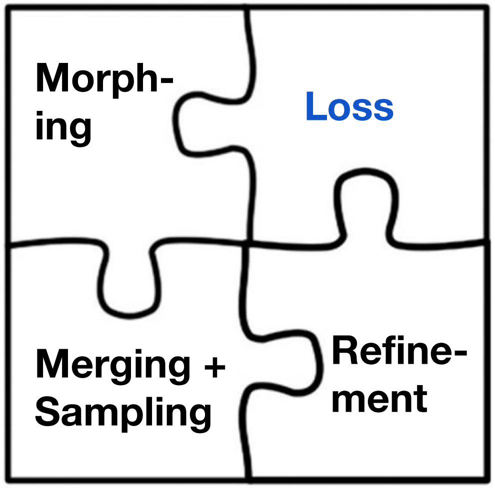


Input
Encoder
GFV
Morphing
-Based
Decoder
Coarse Output
$\mathcal{L}_{expansion}$
Minimum
Density
Sampling
Residual
Network
Point-Wise
Residual
Final Output
MLP$_1$
(1026, 513, 256, 3)
MLP$_K$
(1026, 513, 256, 3)
surface$_1$
surface$_K$
2
1024
GFV
$[0, 1]^2$ Sampling



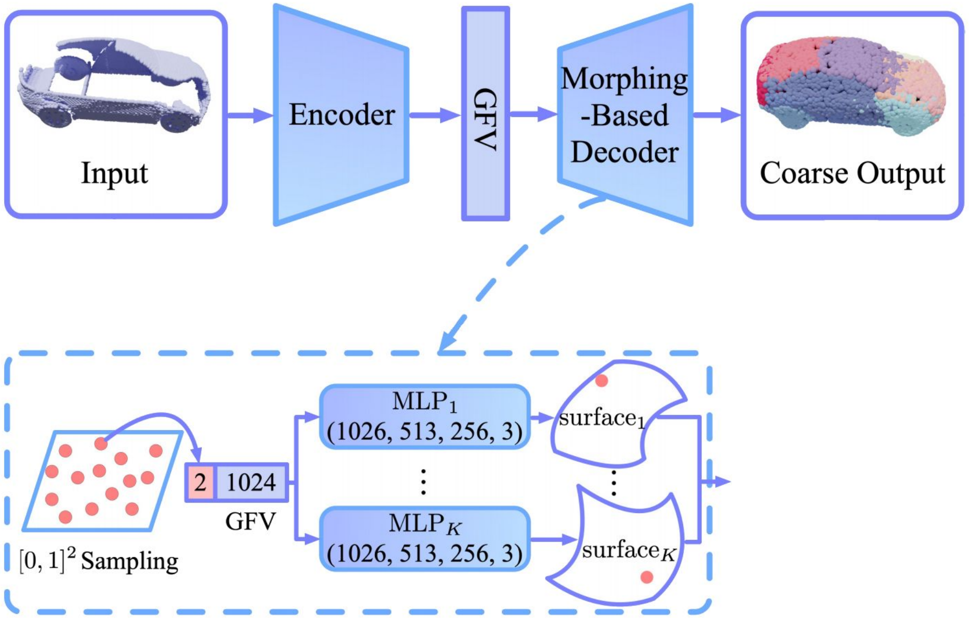
Morph-
ing
Merging +
Sampling
Refine-
ment


Input
Encoder
GFV
Morphing
-Based
Decoder
Coarse Output
$\mathcal{L}_{\text{expansion}}$
$\mathcal{L}_{\text{EMD}}$
Ground Truth
Minimum
Density
Sampling
Residual
Network
Point-Wise
Residual
$\mathcal{L}_{\text{EMD}}$
Final Output
MLP$_1$
(1026, 513, 256, 3)
surface$_1$
2
1024
GFV
MLP$_K$
(1026, 513, 256, 3)
surface$_K$
$[0, 1]^2$ Sampling


Morph-
ing
Loss
Merging +
Sampling
Refine-
ment

1. Encoder and Morphing-Based Decoder

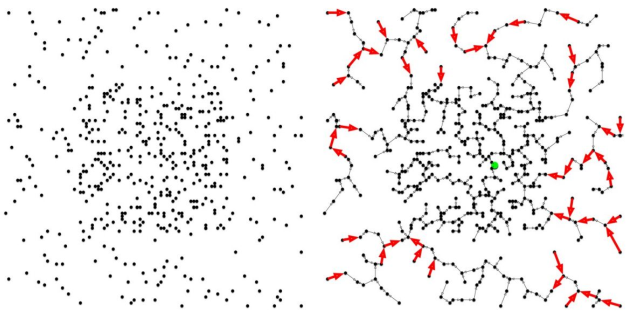
- Encoder
  - Takes point cloud and produces generalized feature vector (based on PointNet)
- Decoder:
  - Start with K 2D grids
  - Sample n points from each and concat with the feature vector
  - Feed them into K MLPs to produce K 3D surface elements (“Morphing”)



1. Expansion Penalty

- Goal: Encourage surface elements to be concentrated in a local area

- Idea: Apply expansion penalty to coarse output
  - Construct directed minimum spanning tree for each surface element
  - Apply loss function to spanning tree:



Point cloud　　Minimum Spanning Tree [1]

$$\mathcal{L}_{\text{expansion}} = \frac{1}{KN} \sum_{1 \leq i \leq K} \sum_{(u,v) \in \mathcal{T}_i} \mathbb{1}\{\text{dis}(u,v) \geq \lambda l_i\} \text{dis}(u,v)$$

Penalizes long edges in the spanning tree
→ Motivates tree to shrink towards center

K: number of surface elements
N: number of points per surface element
T_i: i-th spanning tree
λ: minimum edge length (parameter)

1. Expansion Penalty

Expansion
Penalty
Applied

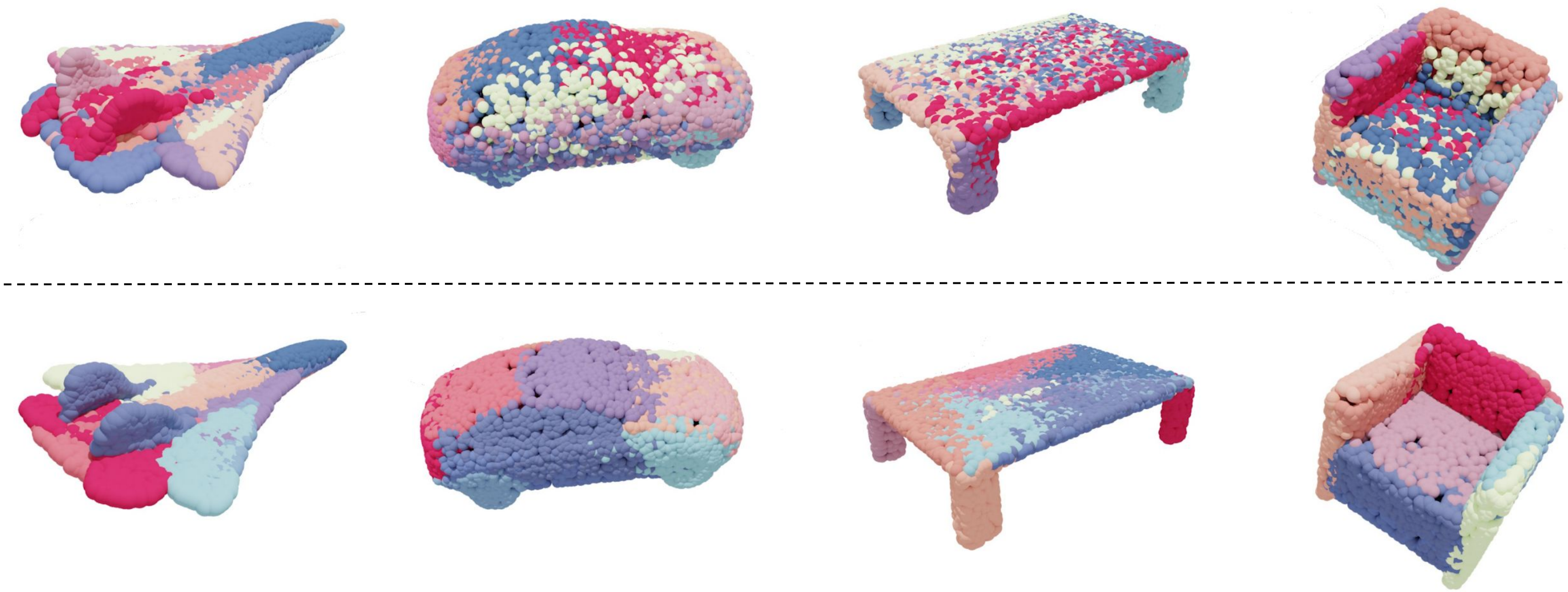


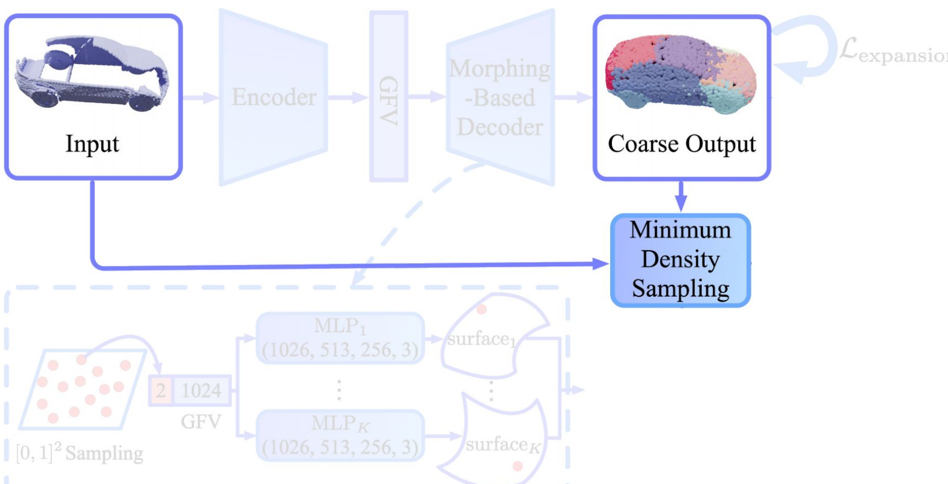
Image adapted from [1]

2. Merging the point clouds

Point cloud so far:

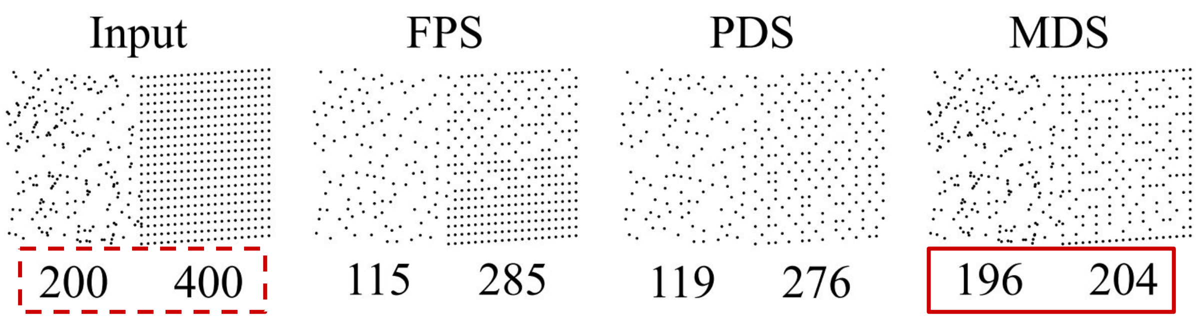
- Smooth surfaces ✓
- Coarse point cloud is evenly distributed ✓
- Fine details of the object not yet modeled
- Some existing structures of the input may have been dropped

  → Merge input point cloud and coarse output



2. Merging the point clouds: Sampling

- Existing Methods
  - Farthest Point Sampling (FPS)
  - Poisson Disk Sampling (PDS)
- Problem:
  - Inputs of different densities -> outputs of different densities
- Solution: Minimum Density Sampling (MDS)
  - Goal: Even density in the output even with varying densities in the input
  - Method: Sample points so that the resulting points cloud has minimum ‘density’



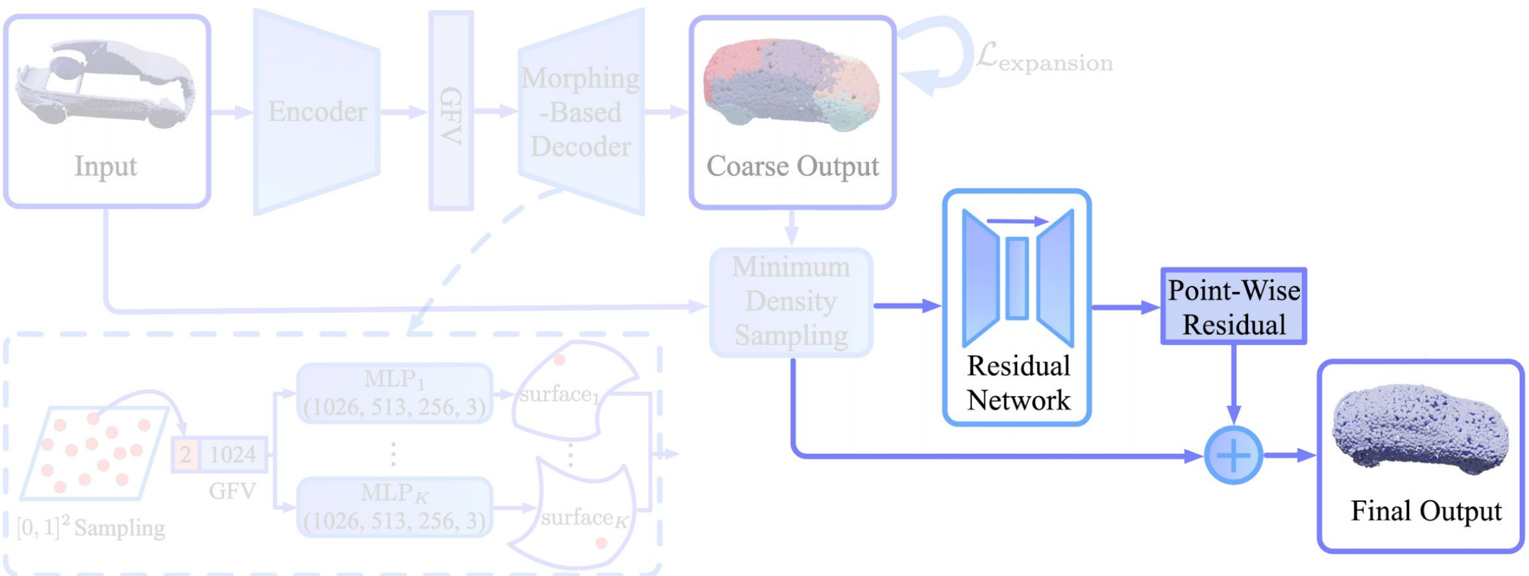
Visualization of sampling algorithms, from [1]

Output of equal density

S

3. Refining

- Residual Network
  - Encoder+Decoder (based on PointNet)
  - Refines details of the point cloud



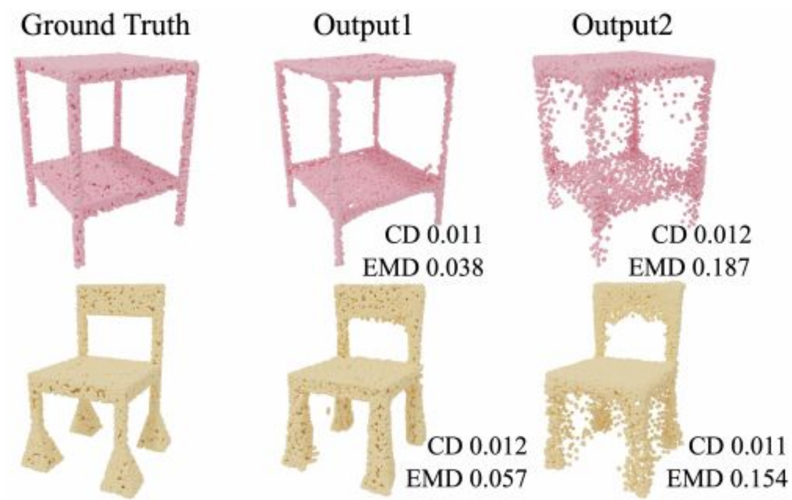
R

4. Loss Function: Distance Metrics

- Chamfer Distance (CD)
  - Distance between each point in a point cloud and its nearest neighbour in the second point cloud
  - Drawbacks:
    - Parts of the output get overpopulated
    - Blurry details

- Earth Mover's Distance (EMD)
  - Find a bijection between two point clouds so that sum of distances between each pair of points is minimized
  - Drawbacks:
    - Point clouds need to be of equal size
    - Memory consumption in $O(n^2)$
      → Infeasible for large point clouds (>2000 points)

→ **Authors Solution: Approximate EMD: Memory in O(n)**



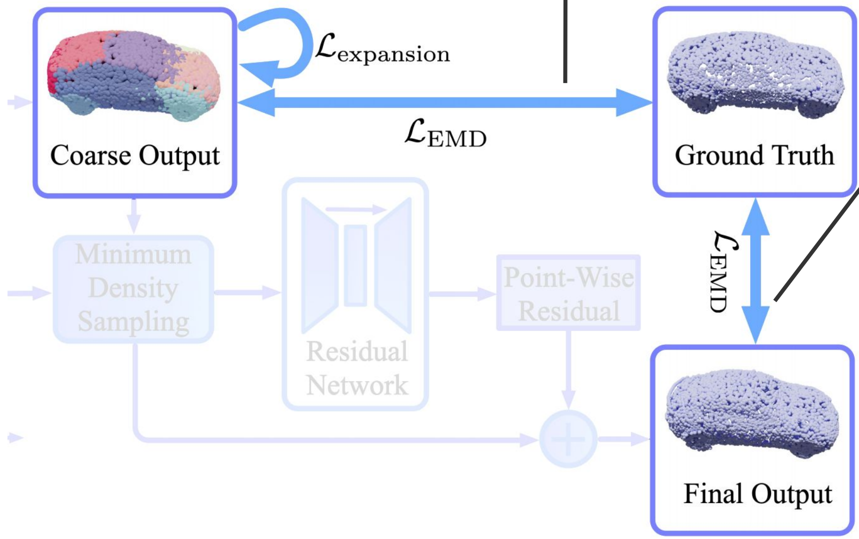
Distance metrics, from [1]

Similar CD,
Much different EMD

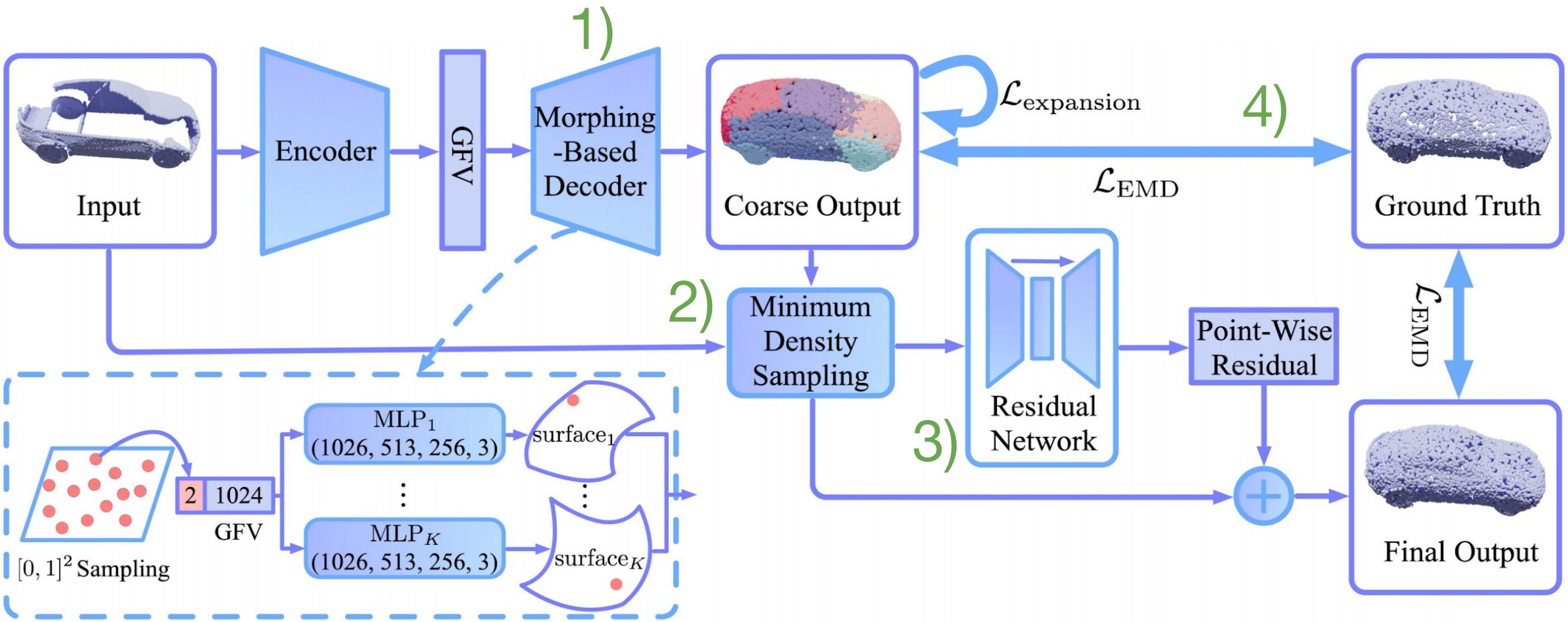
L

4. Loss Function: Final Formula

$$\mathcal{L} = \alpha\mathcal{L}_{\text{expansion}} + \mathcal{L}_{\text{EMD}}(S_{\text{coarse}}, S_{\text{gt}}) + \beta\mathcal{L}_{\text{EMD}}(S_{\text{final}}, S_{\text{gt}})$$



Putting it all together



Our Goals

1. Continuous and smooth surfaces ✓ Morphing-Based Decoder (1)
2. Capture fine details of the object ✓ Refinement (3)
3. Locally even distribution of points ✓ Sampling (2) and EMD+Expansion loss (4)
4. Preserve existing structures from the input ✓ Merging (2)

Evaluation: Quantitative Results

- Evaluated on ShapeNet dataset
  - 8 classes: table, chair, car, airplane, sofa, lamp, vessel, cabinet
  - Uneven distribution of classes (overrepresentation of airplanes and cars)
- EMD and CD distance metrics

methods	vessel	cabinet	table	airplane	car	chair	sofa	lamp	average
Oracle	0.93	1.44	1.21	0.69	1.25	1.17	1.28	0.91	1.11
FCAE	7.22	11.20	7.77	4.10	7.00	7.64	7.00	14.64	8.32
AtlasNet	8.11	8.91	5.07	3.27	4.20	5.03	6.97	10.71	6.53
PCN	6.56	8.79	6.84	3.44	4.44	6.89	6.28	15.45	7.34
Ours	**3.83**	**4.16**	**3.66**	**2.18**	**3.28**	**3.63**	**3.47**	**6.04**	**3.78**

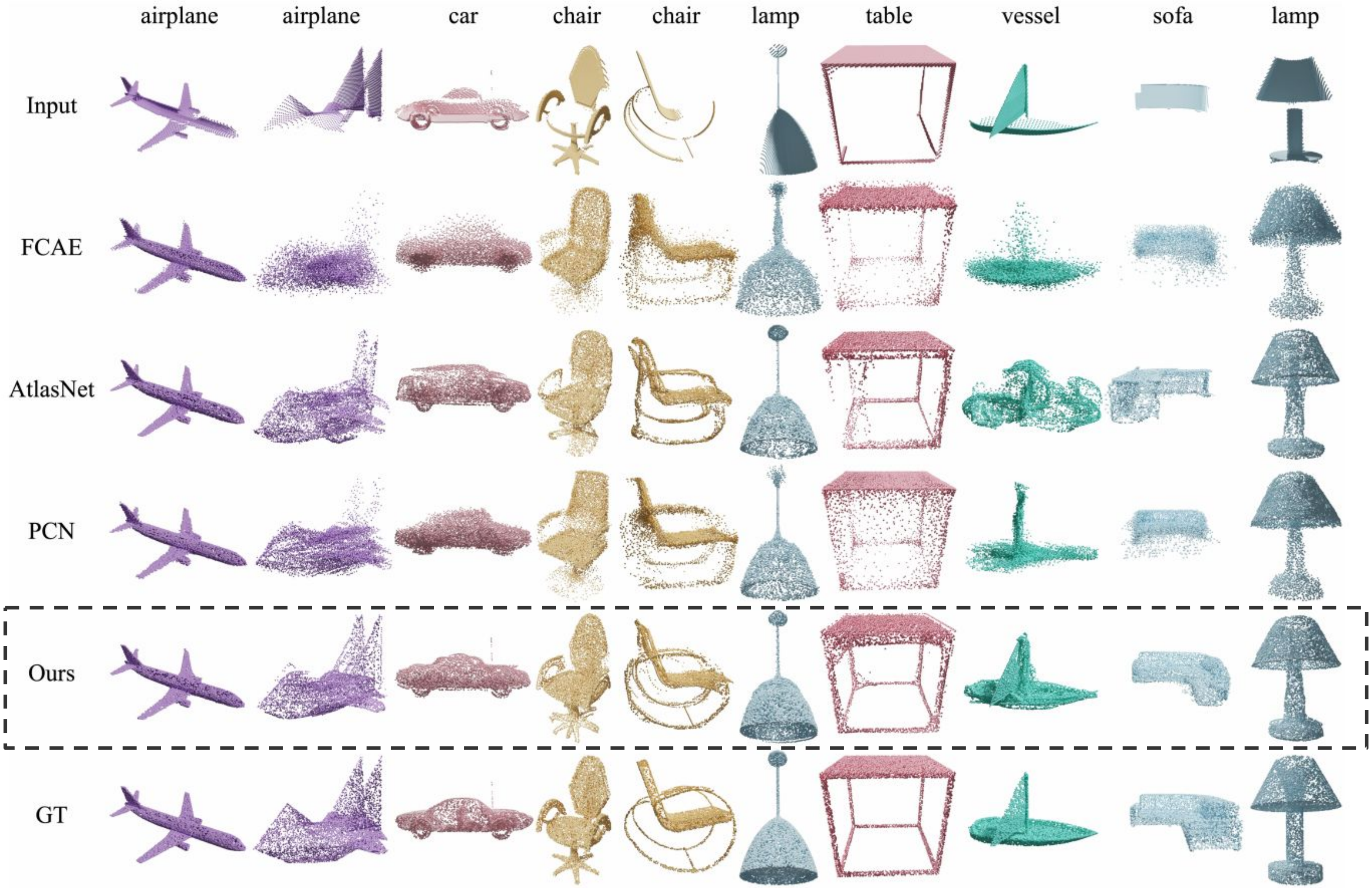
(a) EMD $\times$ 100

methods	vessel	cabinet	table	airplane	car	chair	sofa	lamp	average
Oracle	0.42	0.72	0.55	0.28	0.63	0.50	0.60	0.35	0.50
FCAE	1.33	1.40	1.16	0.70	1.10	1.29	1.37	1.72	1.26
AtlasNet	2.30	2.49	1.46	0.85	1.42	1.58	2.67	1.82	1.82
PCN	1.23	1.35	1.14	0.66	1.10	1.41	1.36	1.46	1.21
Ours	1.17	1.37	1.15	0.60	1.11	1.16	1.31	1.30	1.14
Ours-CD	**0.99**	**1.19**	**0.96**	**0.56**	**1.03**	**1.02**	**1.16**	**1.07**	**1.00**

(b) CD $\times$ 100

Quantitative Results of MSN in relation to SOTA, from [1]

Evaluation: Qualitative Results



Qualitative Results of MSN in relation to SOTA, from [1]

Conclusion

- Contributions:
  - A two-stage approach for point cloud completion: from coarse to fine
  - An expansion penalty to control the distribution of the surface elements
  - A novel sampling algorithm for point clouds with evenly distributed results
  - An approximation of the Earth Mover's Distance

- Performance:
  - Net beats SOTA in both distance metrics (CD and EMD)

- Limitations:
  - MDS preserves clutter in the input cloud

My take on the paper

Method:

- Extensive ablation studies
- Evaluation on only one dataset

Architecture:

- Possible improvements:
  - Combine folding-based and fully connected decoder (like in GRNet)
  - Explicitly predict missing parts

General assessment:

- Code available
- Network architecture is well explained
- Authors have presented solutions to existing problems which can prove useful for further research:
  - EMD approximation
  - MDS sampling algorithm
- Recent advances in voxel-based methods show promising results (see GRNet) - Point-cloud based methods still the way to go?

Sources

[1] MSN:
Liu, M., Sheng, L., Yang, S., Shao, J., & Hu, S. M. (2020, April). Morphing and sampling network for dense point cloud completion. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 11596-11603).
[2] PC Completion:
cs.cmu.edu/~wyuan1/pcn/images/shapenet.png, last accessed 12.01.2021
[3] Symmetry:
Thrun, S., & Wegbreit, B. (2005, October). Shape from symmetry. In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 (Vol. 2, pp. 1824-1831). IEEE.
[4] Example DB:
Pauly, M., Mitra, N. J., Giesen, J., Gross, M. H., & Guibas, L. J. (2005). Example-based 3D scan completion. In Symposium on Geometry Processing (No. CONF, pp. 23-32).
[5] Point Clouds:
Lim, I., Ibing, M., & Kobbelt, L. (2019, August). A Convolutional Decoder for Point Clouds using Adaptive Instance Normalization. In Computer Graphics Forum (Vol. 38, No. 5, pp. 99-108).
[6] PointNet:
Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 652-660).
[7] FoldingNet:
Yang, Y., Feng, C., Shen, Y., & Tian, D. (2018). Foldingnet: Point cloud auto-encoder via deep grid deformation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 206-215).
[8] PCN:
Yuan, W., Khot, T., Held, D., Mertz, C., & Hebert, M. (2018, September). Pcn: Point completion network. In 2018 International Conference on 3D Vision (3DV) (pp. 728-737). IEEE.
[9] GRNet:
Xie, H., Yao, H., Zhou, S., Mao, J., Zhang, S., & Sun, W. (2020). GRNet: Gridding Residual Network for Dense Point Cloud Completion. arXiv preprint arXiv:2006.03761.
[10] ShapeNet: Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., ... & Yu, F. (2015). Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012.

Any Questions?

Backup

Backup: Minimum Density Sampling

$$p_i = \underset{x \notin \mathrm{P}_{i-1}}{\operatorname{argmin}} \sum_{p_j \in \mathrm{P}_{i-1}} \exp(-\|x - p_j\|^2/(2\sigma^2))$$

P_i: set of previously sampled points $P_i = \{p_i \,|1 \leq j \leq i\}$
σ: size of the neighbourhood considered (parameter)

Backup: Distance Metrics

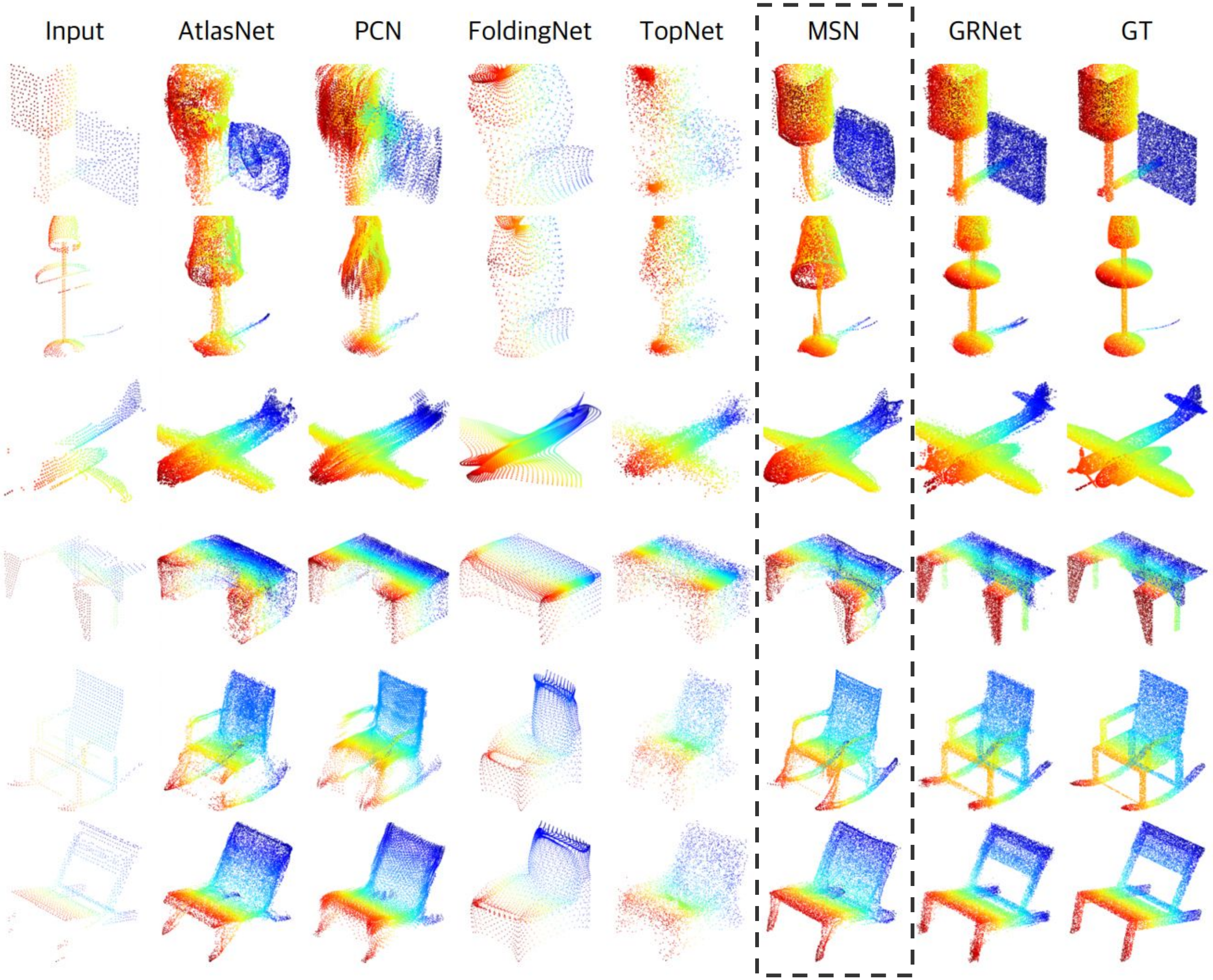
Chamfer Distance (CD)

$$\mathcal{L}_{\mathrm{CD}}(S_1,S_2)=\frac{1}{2}\left(\frac{1}{|S_1|}\sum_{x\in S_1}\min_{y\in S_2}\|x-y\|+\frac{1}{|S_2|}\sum_{y\in S_2}\min_{x\in S_1}\|x-y\|\right)$$

- Earth Mover's Distance (EMD)

$$\mathcal{L}_{\mathrm{EMD}}\left(S_1, S_2\right)=\min_{\phi: S_1 \rightarrow S_2} \frac{1}{\left|S_1\right|} \sum_{x \in S_1}\|x-\phi(x)\|_2$$

Backup: Evaluation - Qualitative Results



Qualitative Results of GRNet [9]

Ablation Studies

- Expansion Loss, Merging (B) and Refining (D):
  - increased CD and EMD loss
- Sampling (C):
  - EMD: best results with MDS
  - CD: best results with FPS
  - Reason: FPS may preserve points from the reliable input at the cost of more uneven distribution

methods	vessel	cabinet	table	airplane	car	chair	sofa	lamp	average
A	3.94	4.33	3.85	2.23	3.47	3.78	3.59	6.08	3.91
B	4.18	4.37	4.08	2.39	3.46	3.89	3.75	6.51	4.08
C	4.30	5.30	4.24	2.59	4.01	4.41	4.18	6.38	4.43
D	3.93	4.32	3.73	2.38	3.41	3.73	3.64	6.02	3.89
E	5.44	6.81	4.52	3.01	4.39	5.44	5.62	8.93	5.52
Ours	**3.83**	**4.16**	**3.66**	**2.18**	**3.28**	**3.63**	**3.47**	**6.04**	**3.78**

(a) EMD $\times$ 100

methods	vessel	cabinet	table	airplane	car	chair	sofa	lamp	average
A	1.20	1.46	1.22	0.62	1.15	1.23	1.38	1.37	1.20
B	1.36	1.48	1.29	0.70	1.19	1.30	1.45	1.59	1.29
C	1.09	1.38	1.12	0.58	1.11	1.10	1.27	1.23	1.11
D	1.24	1.44	1.21	0.64	1.15	1.23	1.40	1.39	1.21
E	**0.99**	**1.19**	**0.96**	**0.56**	**1.03**	**1.02**	**1.16**	**1.07**	**1.00**
Ours	1.17	1.37	1.15	0.60	1.11	1.16	1.31	1.30	1.14

(b) CD $\times$ 100

methods	$\mathcal{L}_{\text{expansion}}$	merging	refining	CD/EMD
w/o $\mathcal{L}_{\text{expansion}}$ (A)	$\times$	MDS	✓	EMD
w/o merging (B)	✓	$\times$	$\times$	EMD
w/o MDS (C)	✓	FPS	✓	EMD
w/o refining (D)	✓	MDS	$\times$	EMD
Ours-CD (E)	✓	MDS	✓	CD
Ours	✓	MDS	✓	EMD

Qualitative results of the ablated net, from [1]