Img 2 Mesh

Estevam Fernandes Arantes 9763105
João Victor Almeida de Aguiar 8503986

Goal

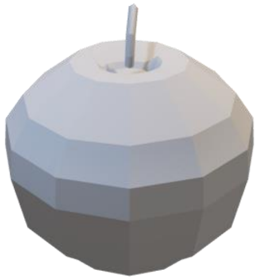
Regression

$\vec{X}$

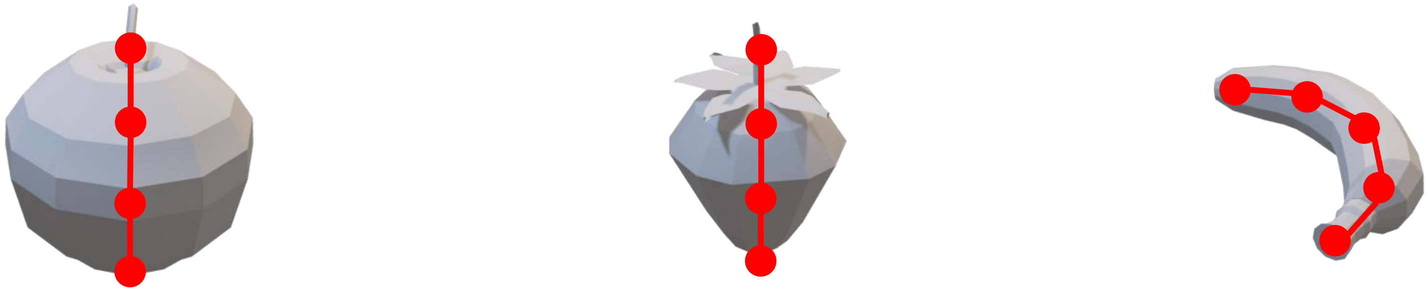
$\vec{Y}$

$\hat{f}(\vec{X})$

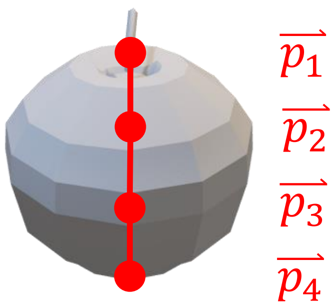
Synthetic Dataset



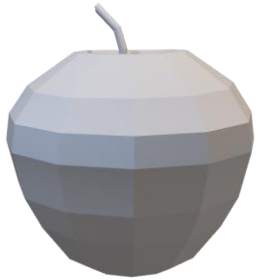
Synthetic Dataset



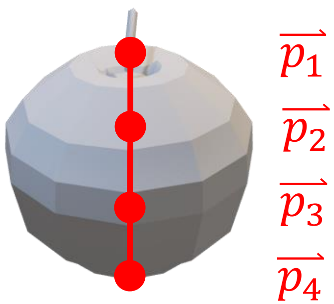
Synthetic Dataset



$$\overrightarrow{p_i} = random(min_i, max_i)$$

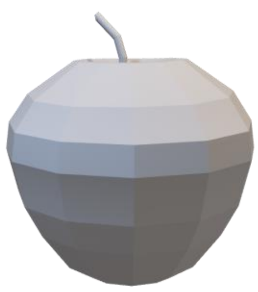


Synthetic Dataset

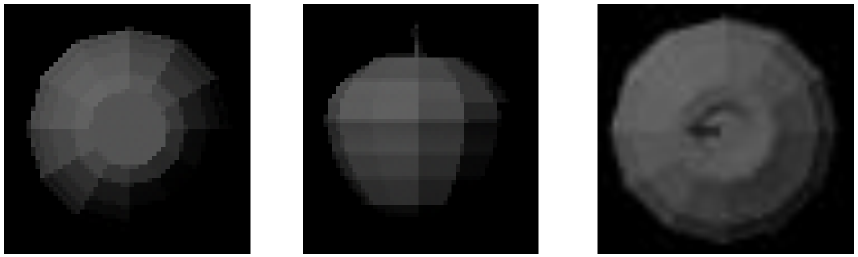


$$\overrightarrow{p_i} = random(min_i, max_i)$$

$\vec{Y}$

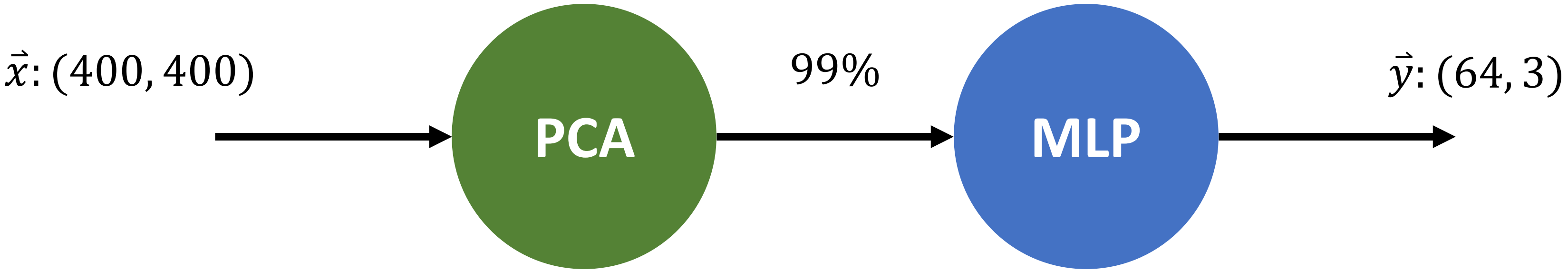


$render$

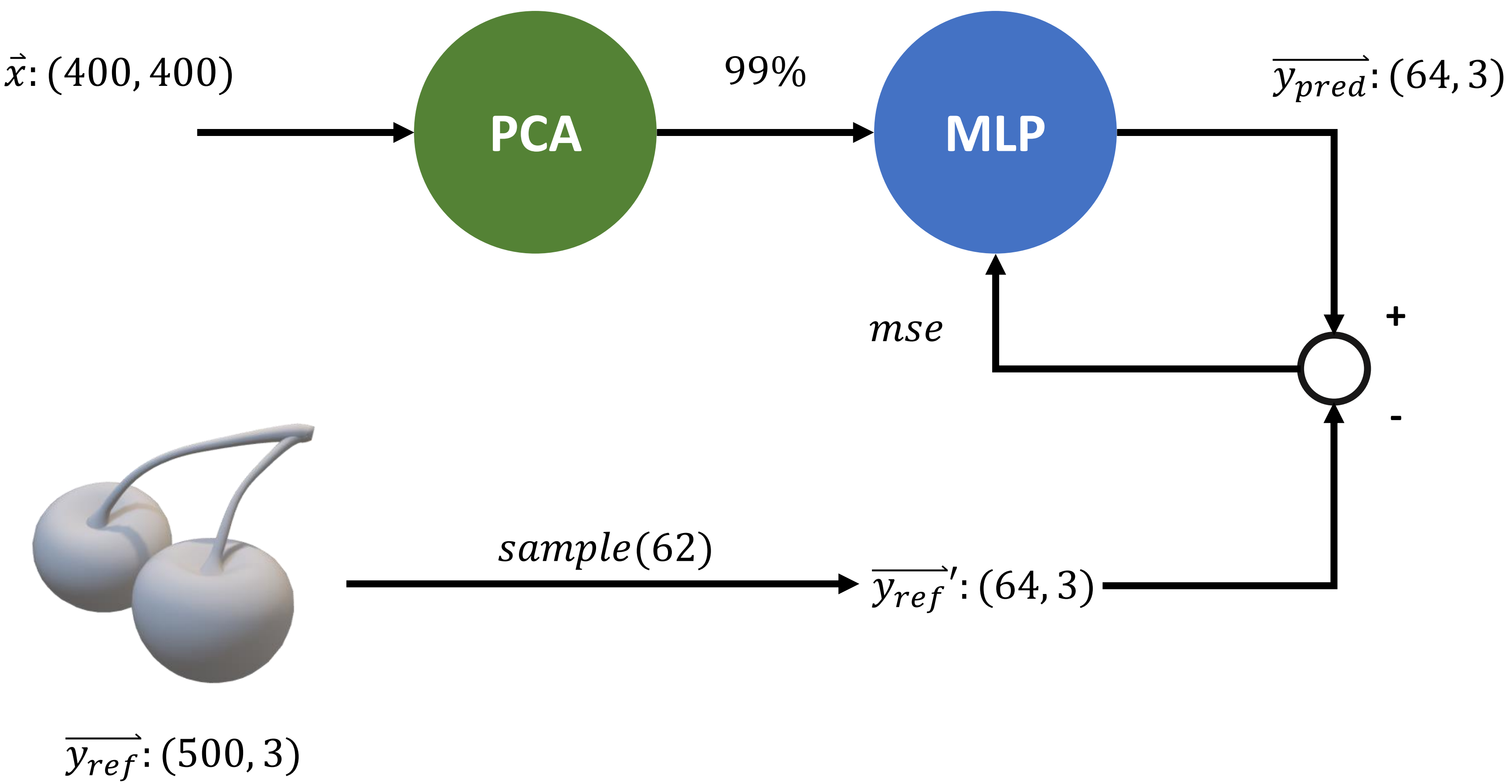


$\vec{X}$

Naïve Solution

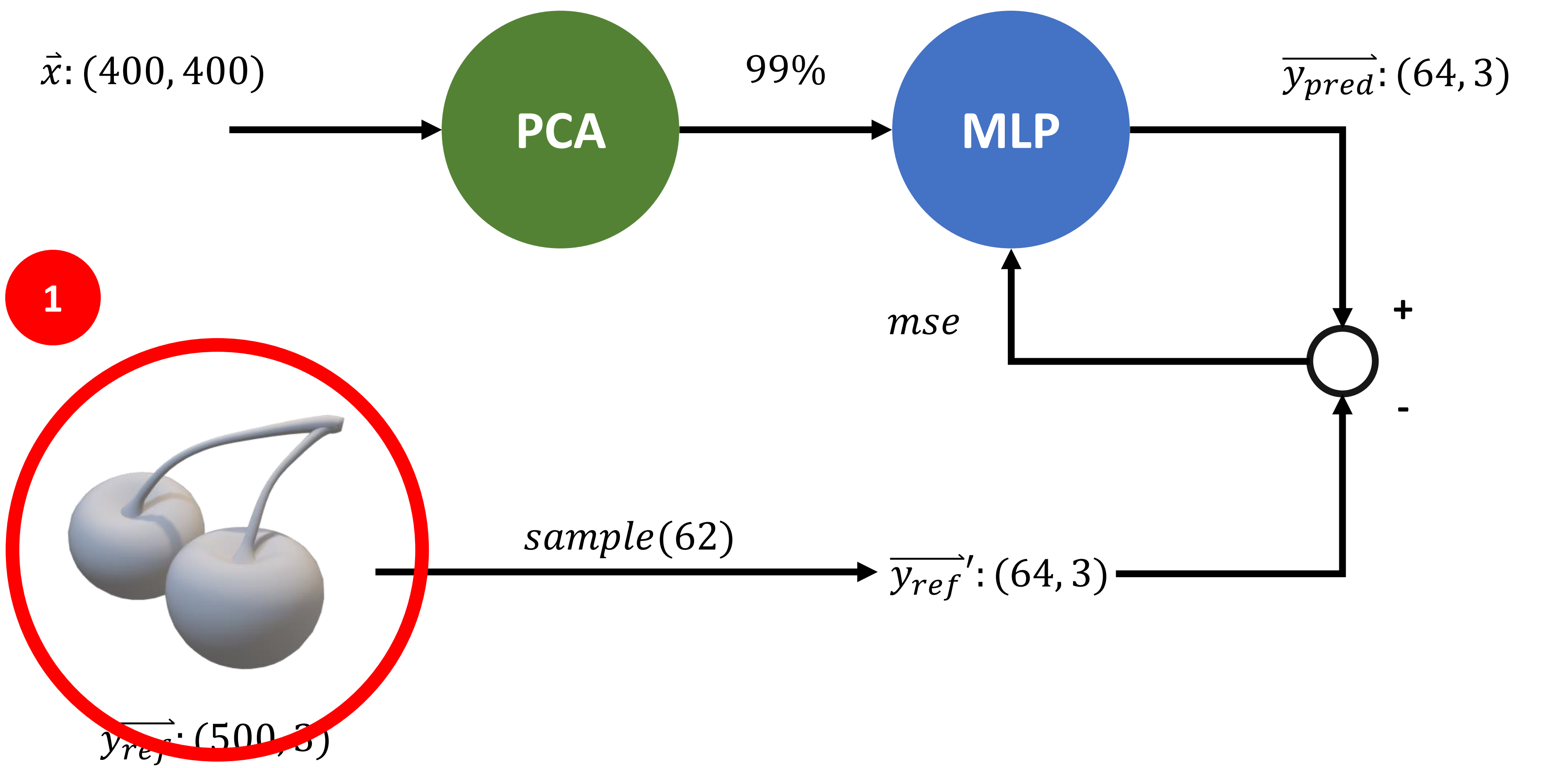


Naïve Solution (Training)



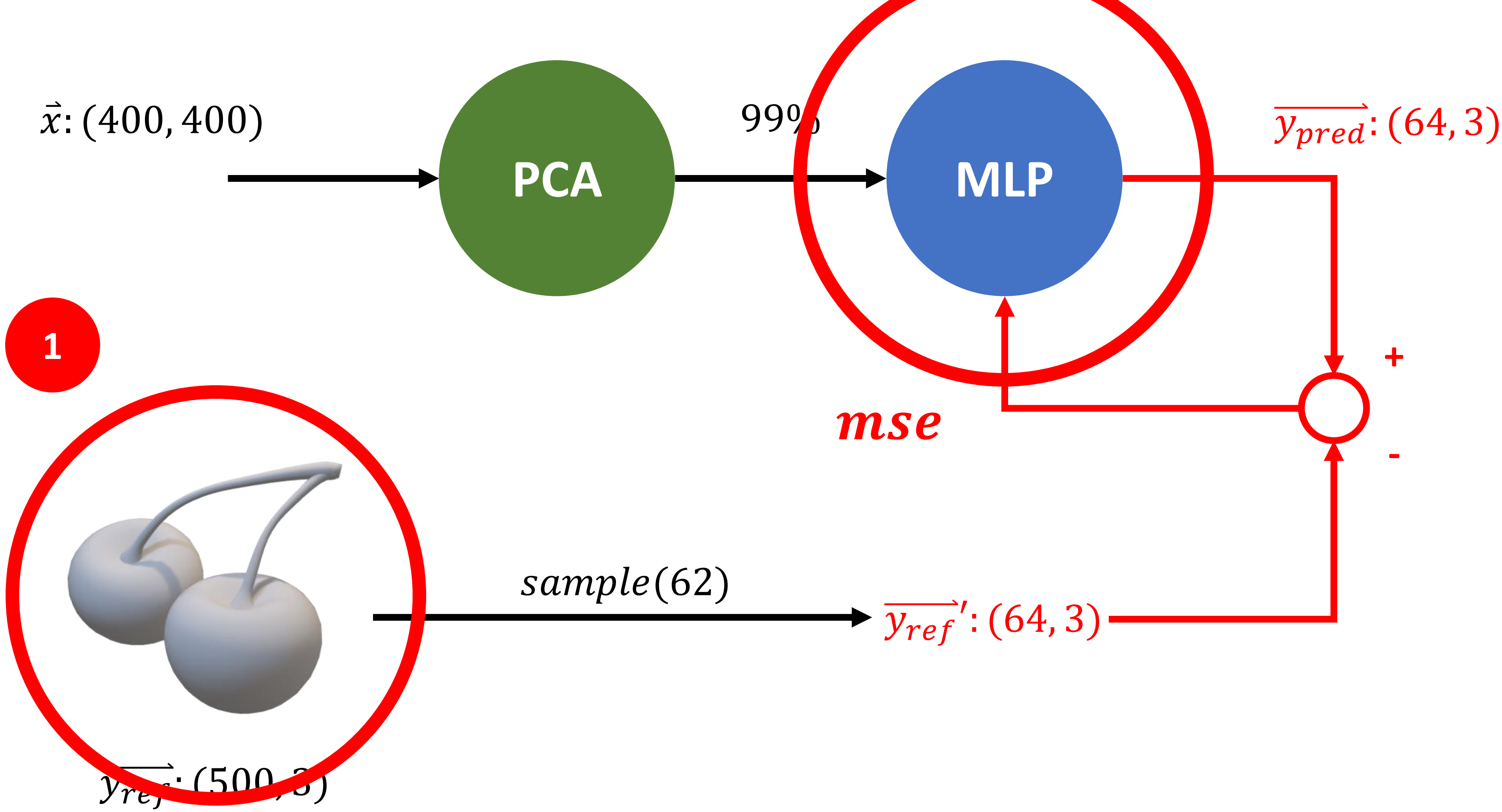
Problem 1 : Dataset

Too biased dataset



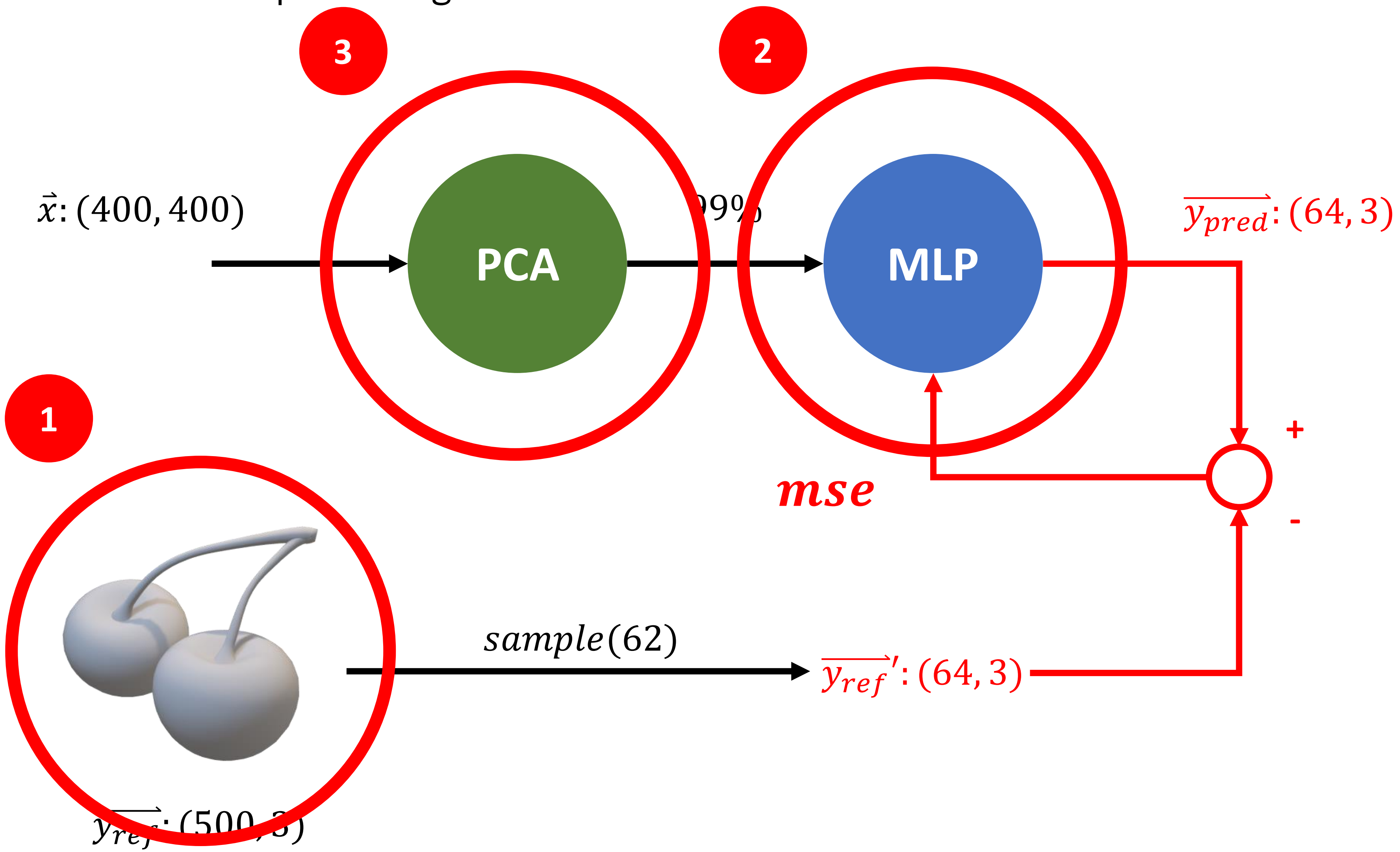
Problem 2 : Loss Metric

mse
- can't compare sets of different sizes
- compares vertex 'i' with vertex 'i'



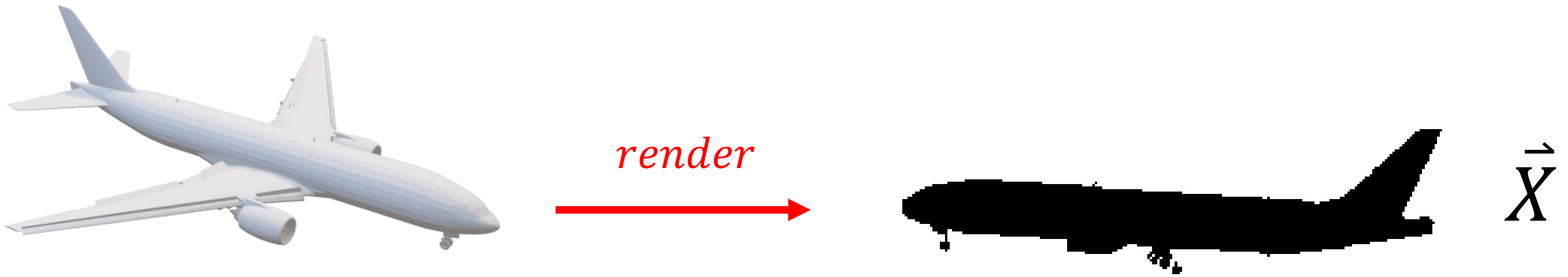
Problem 3 : Encoding

PCA don’t use pixel neighborhood information



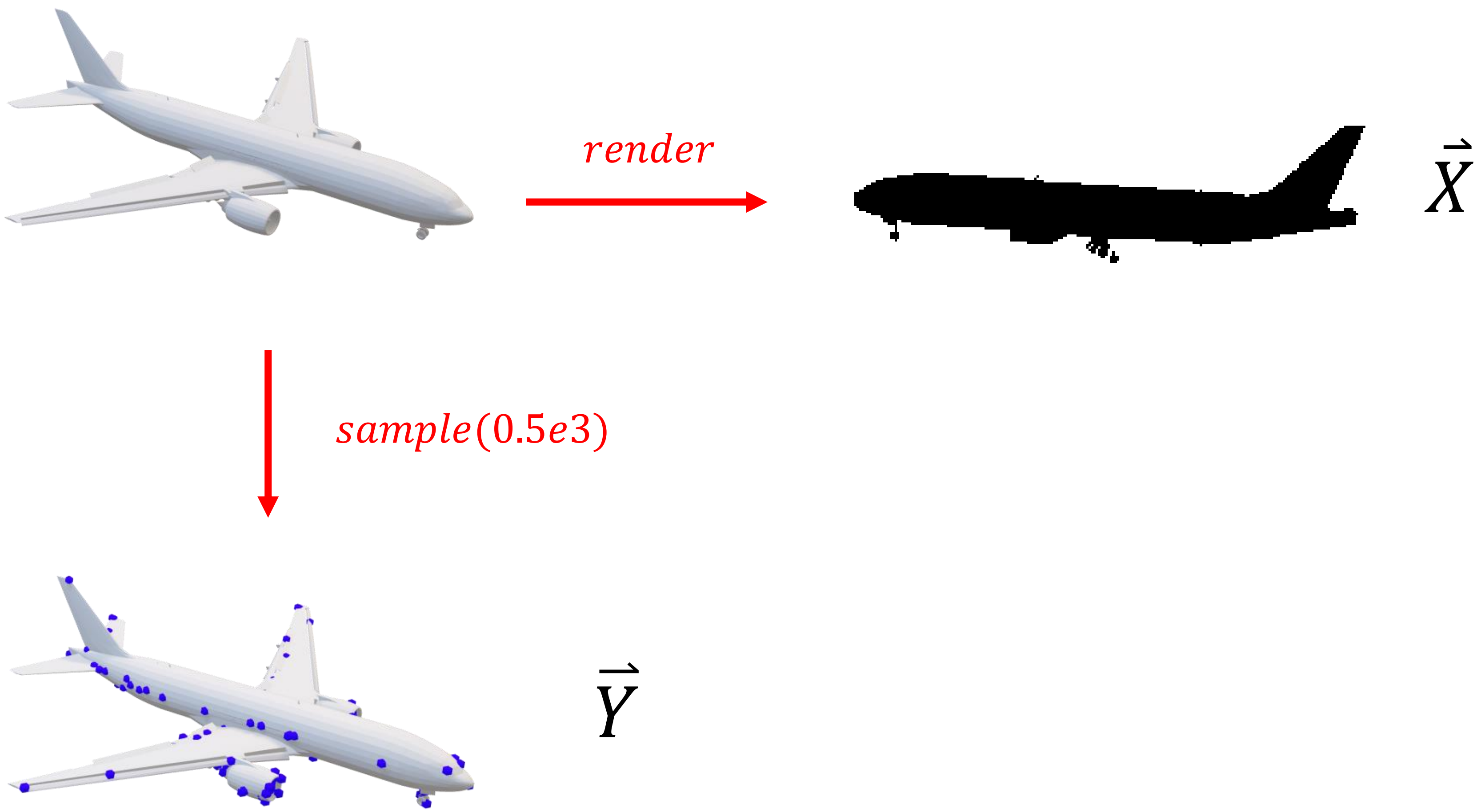
Solution 1 : ModelNet Dataset

Each object is rendered with fixed cameras



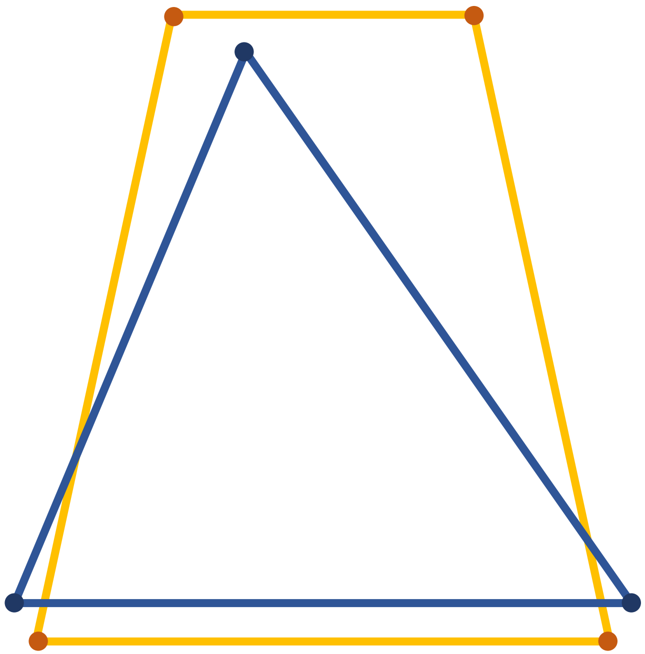
Solution 1 : ModelNet Dataset

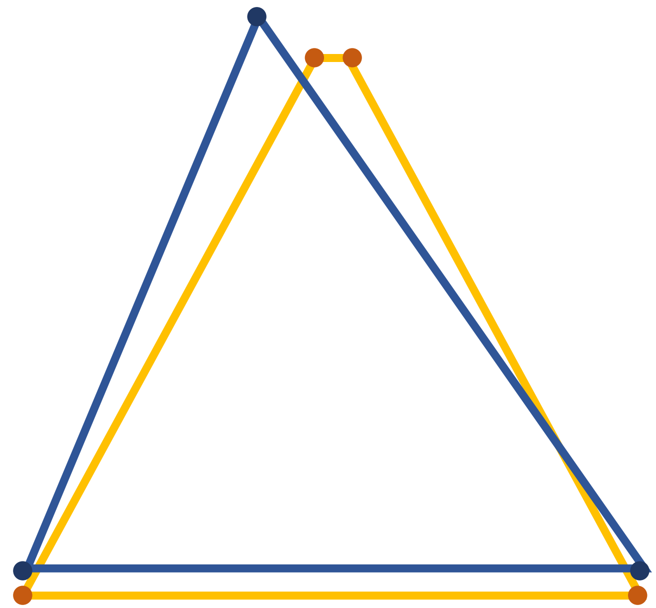
Models are downsampled for faster training



Solution 2 : Chamfer Loss

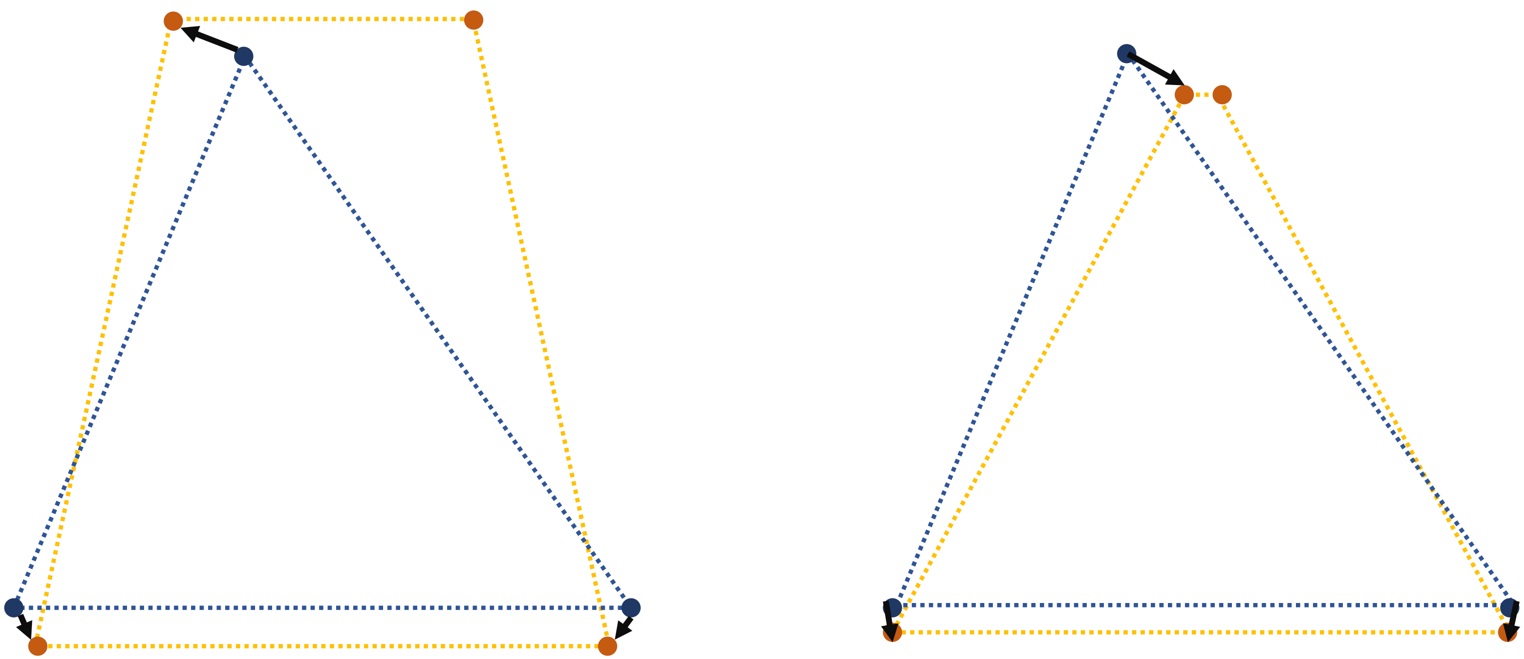
- Vertices order independent
- Compares != number of vertices





$$\overrightarrow{y_{pred}} \; vs \; \overrightarrow{y_{real\,1}} \quad > \quad \overrightarrow{y_{pred}} \; vs \; \overrightarrow{y_{real\,2}}$$

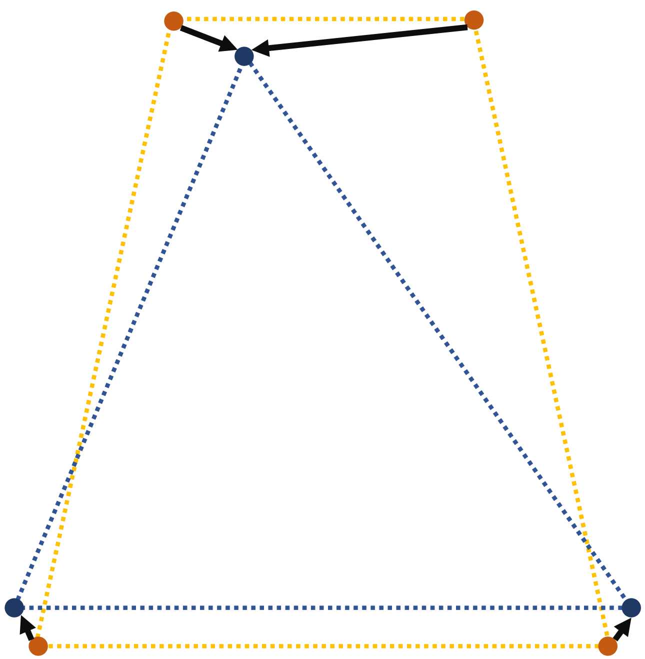
Solution 2 : Chamfer Loss (Pred)



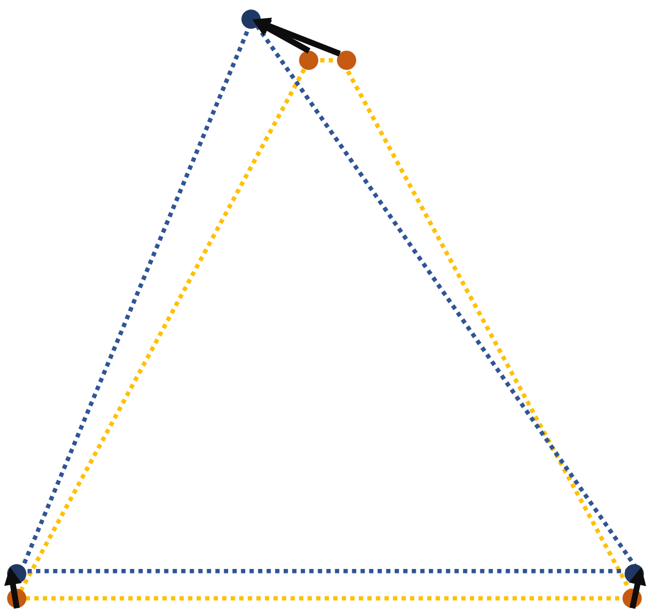
$\overrightarrow{y_{pred}}$ vs $\overrightarrow{y_{real\,1}}$

$\overrightarrow{y_{pred}}$ vs $\overrightarrow{y_{real\,2}}$

Solution 2 : Chamfer Loss (Real)

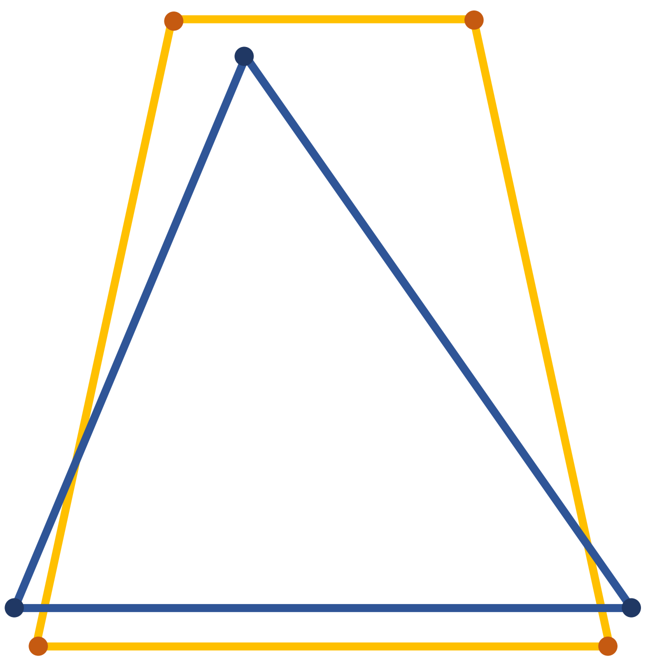


$\overrightarrow{y_{pred}}$ vs $\overrightarrow{y_{real\,1}}$



$\overrightarrow{y_{pred}}$ vs $\overrightarrow{y_{real\,2}}$

Solution 2 : Chamfer Loss



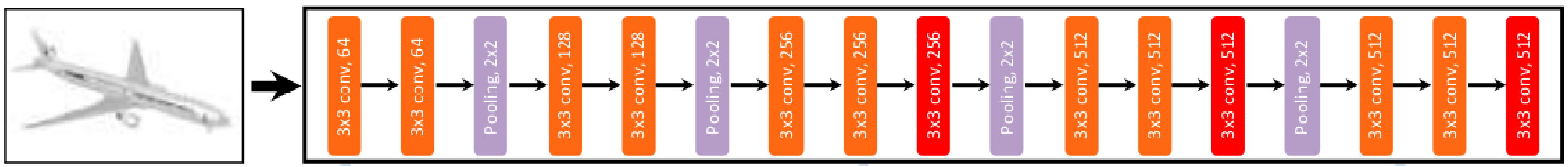
$\overrightarrow{y_{pred}}$ *vs* $\overrightarrow{y_{real\,1}}$

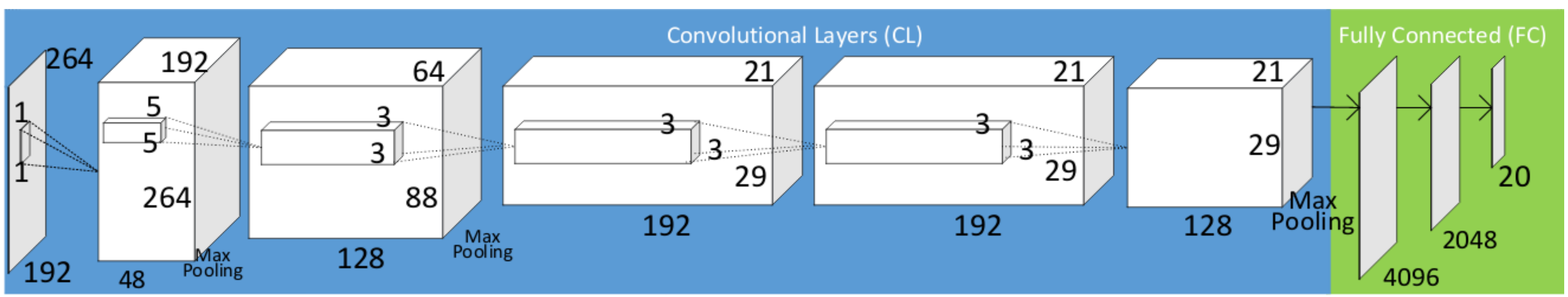
$$loss(p,r) = 0,5 \cdot \frac{1}{N_p} \cdot \sum_{i=1}^{N_p} \min_j \left|\overrightarrow{p_i} - \overrightarrow{q_j}\right|$$

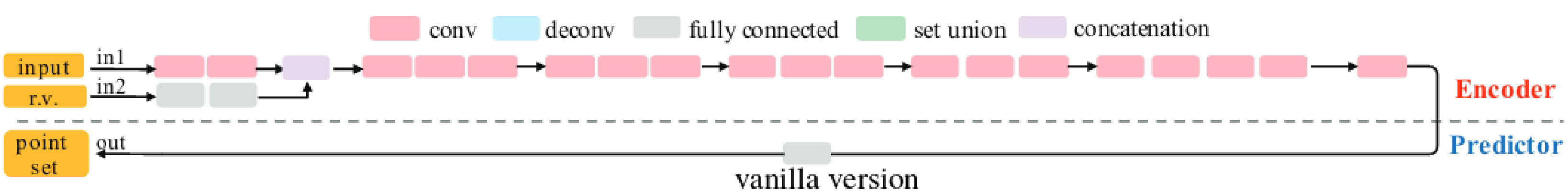
$$+ 0,5 \cdot \frac{1}{N_q} \cdot \sum_{j=1}^{N_q} \min_i \left|\overrightarrow{p_i} - \overrightarrow{q_j}\right|$$

Solution 3: Deep Learning

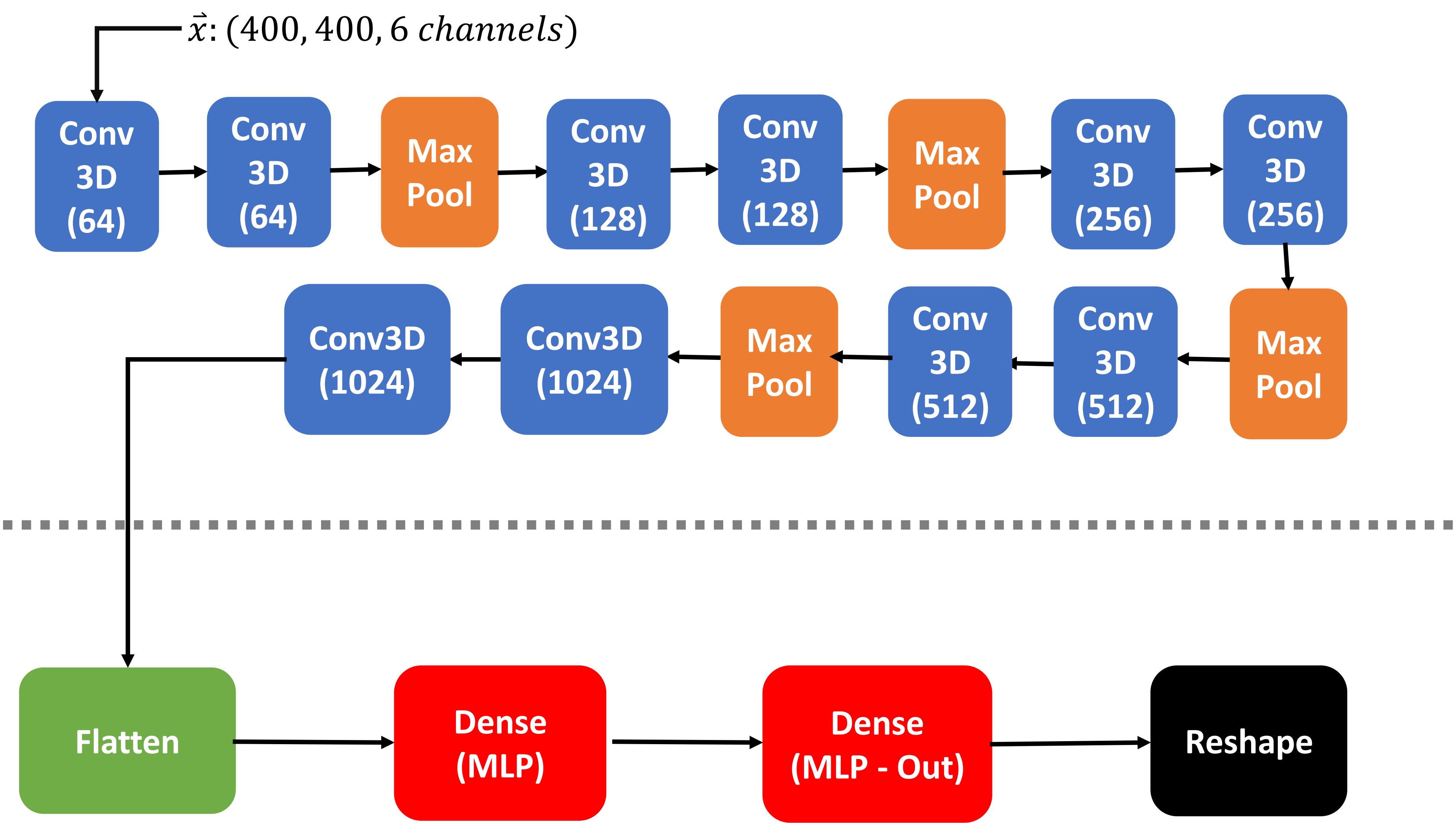
- PCA - Linear
- Most articles - CNNs [1, 2, 3]





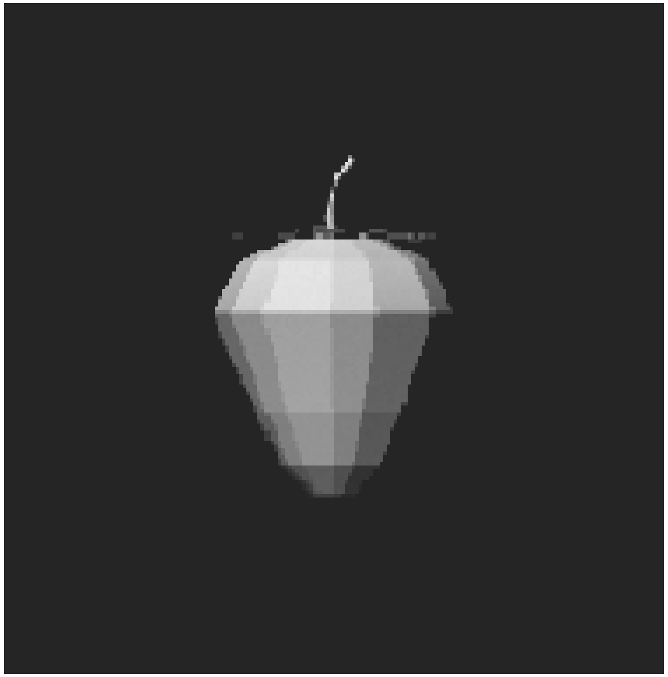


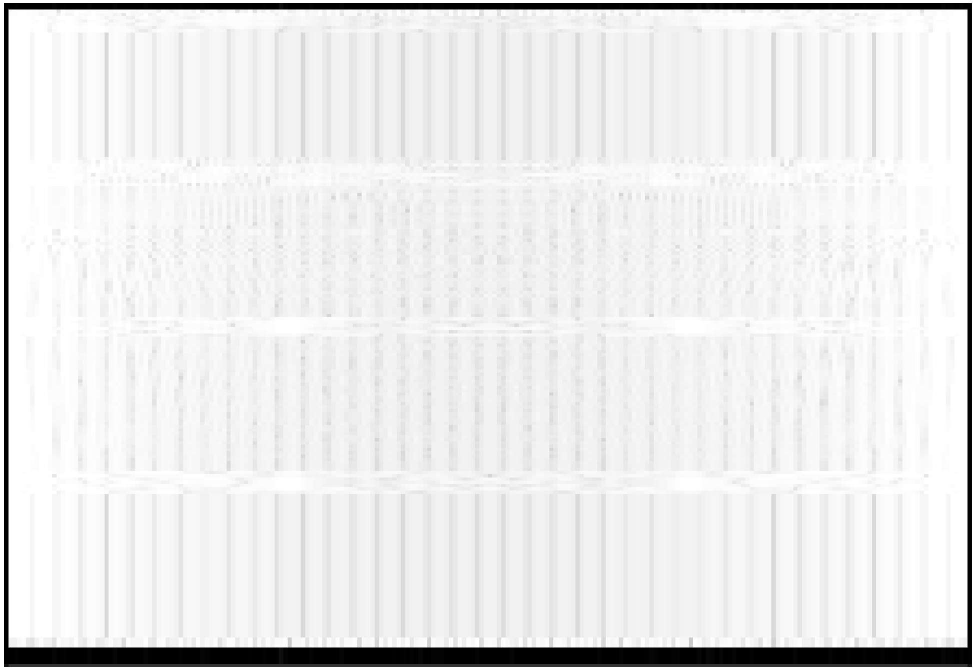
VGG Architecture (insipired)



VGG Architecture : Filtering

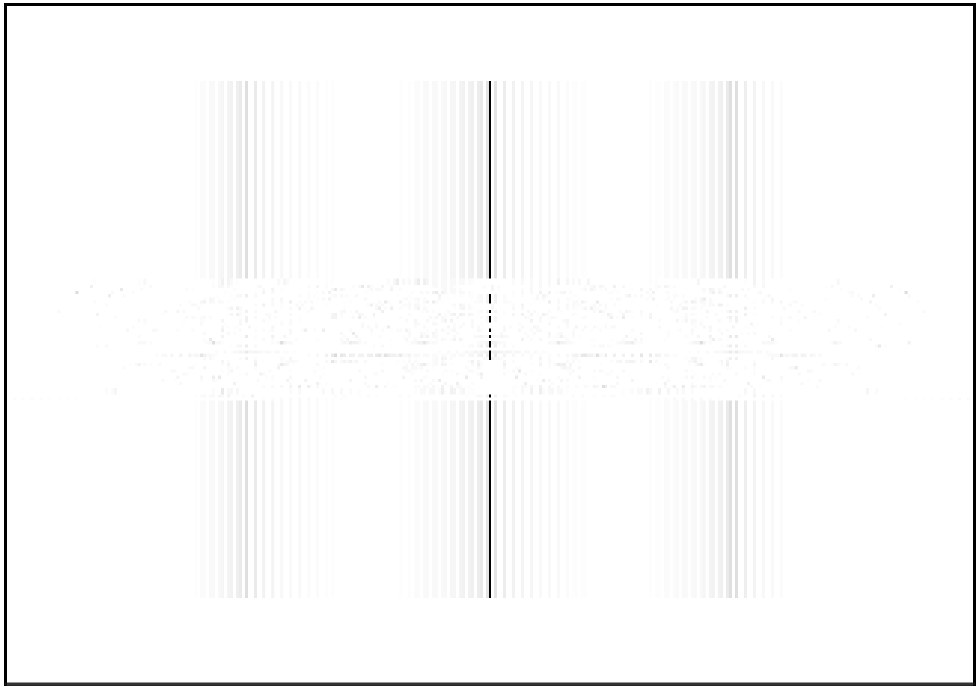
Most articles : render with lighthing => power spectrum in many frequencies



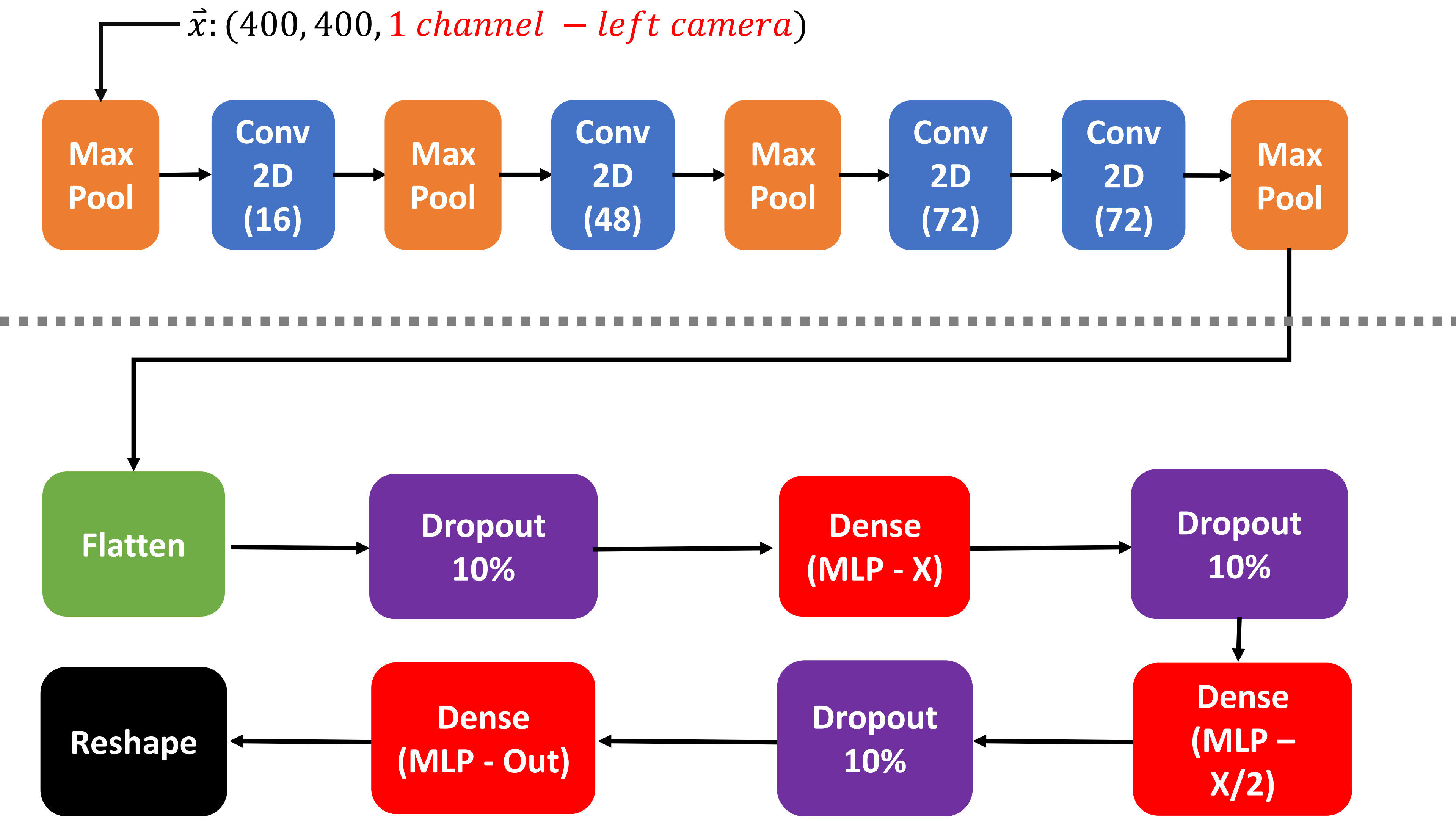


VGG Architecture : Filtering

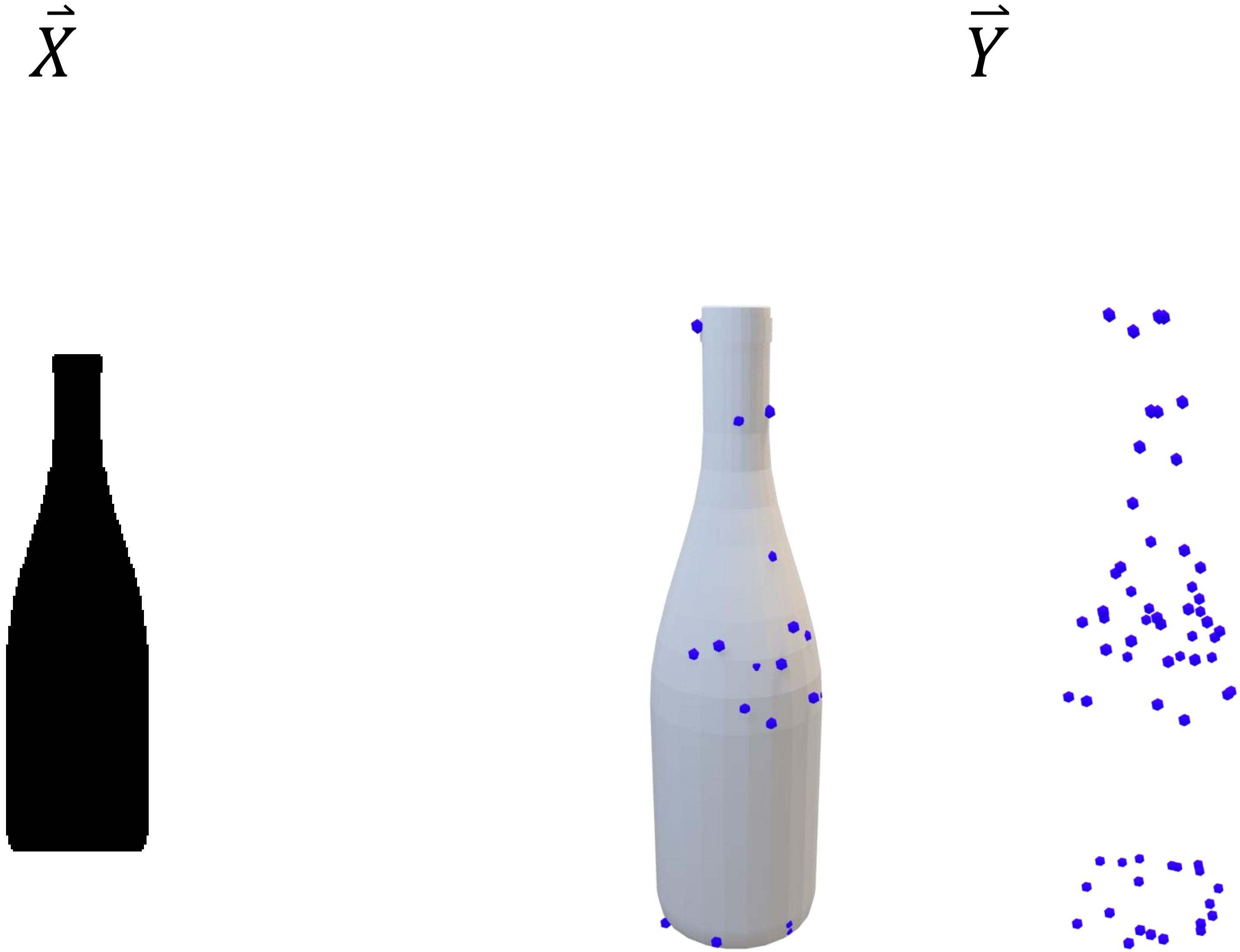
Our solution: silhouette => high frequency information
Less CNN / MaxPool layers may give a better result



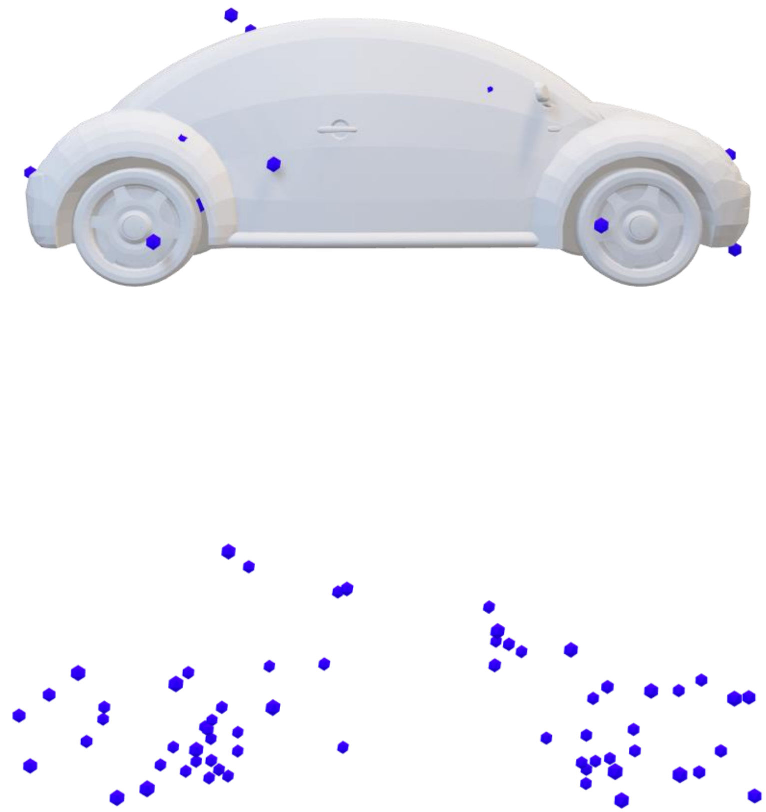
Silhouette Architecture



$\vec{X}$ $\vec{Y}$

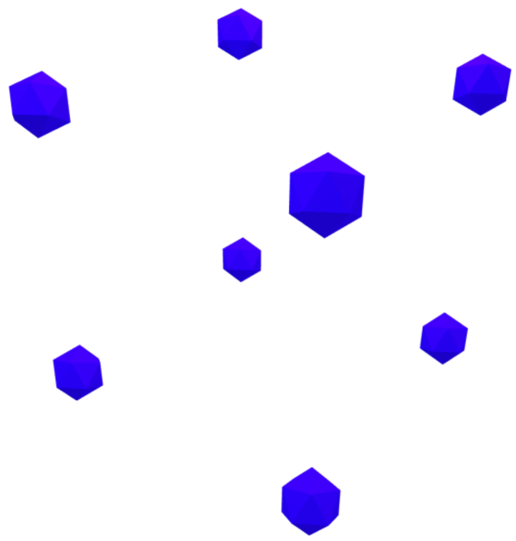

$\vec{X}$
$\vec{Y}$

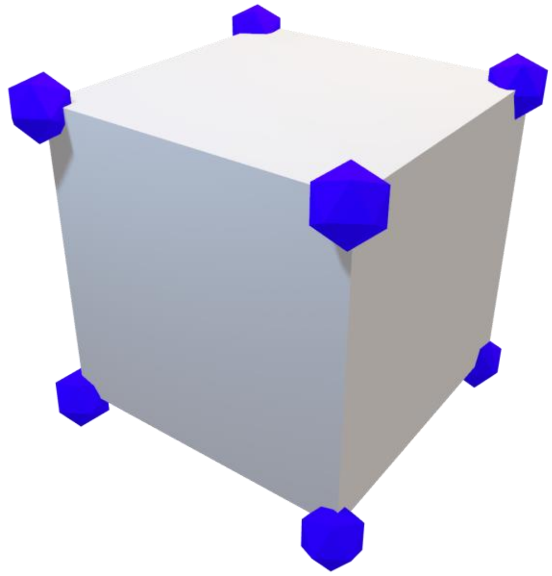
$\vec{X}$ $\vec{Y}$



Mesh ?

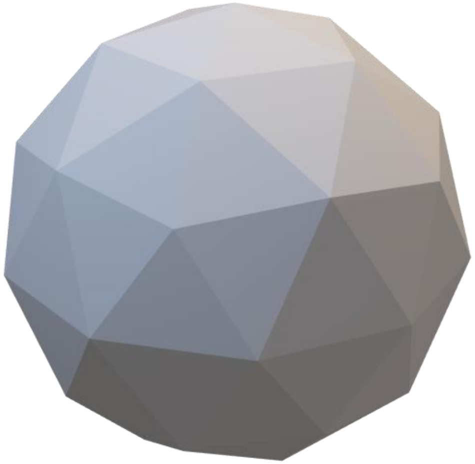
$\overrightarrow{y_{pred}}: (64, 3) \longrightarrow \{vertices, faces\}$





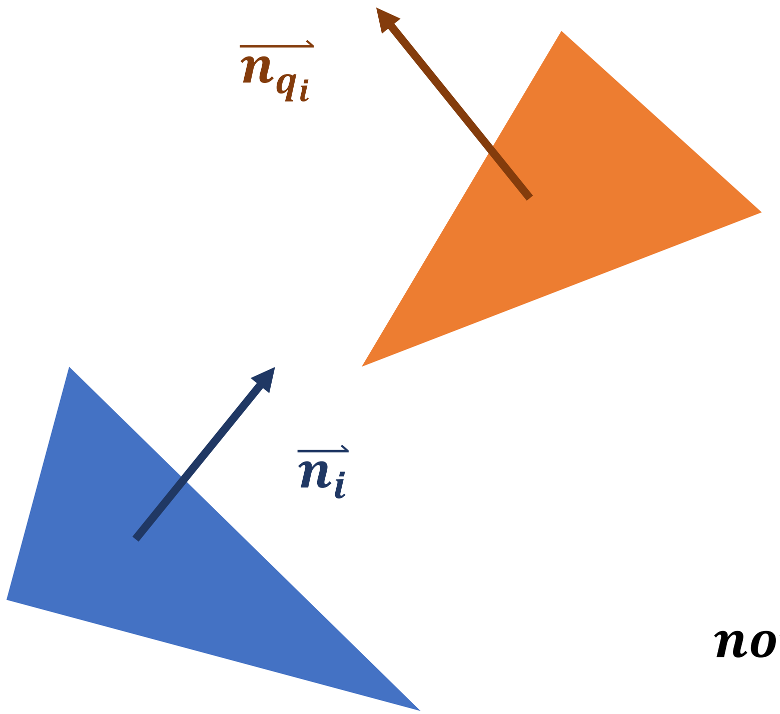
Mapping!

$\overrightarrow{y_{pred}}: (64, 3)$ → $\{vertices = y_{pred}\}$



→ $\{faces\}$

Normal Loss

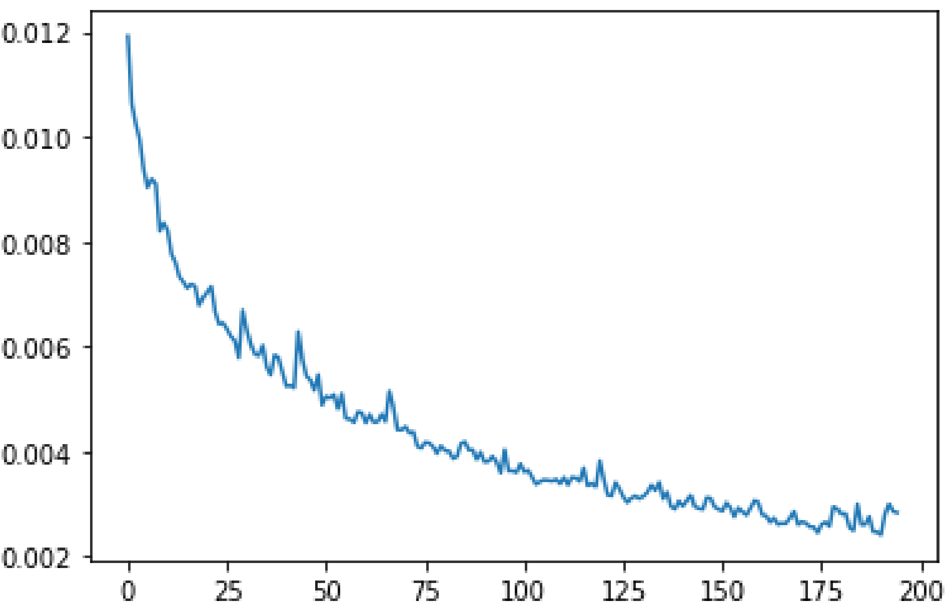


$$normal\_loss(p, q) = 1 - \frac{1}{N_p} \cdot \sum_{i=1}^{N_p} < \overrightarrow{n_i}, \overrightarrow{n_{q_i}} >$$

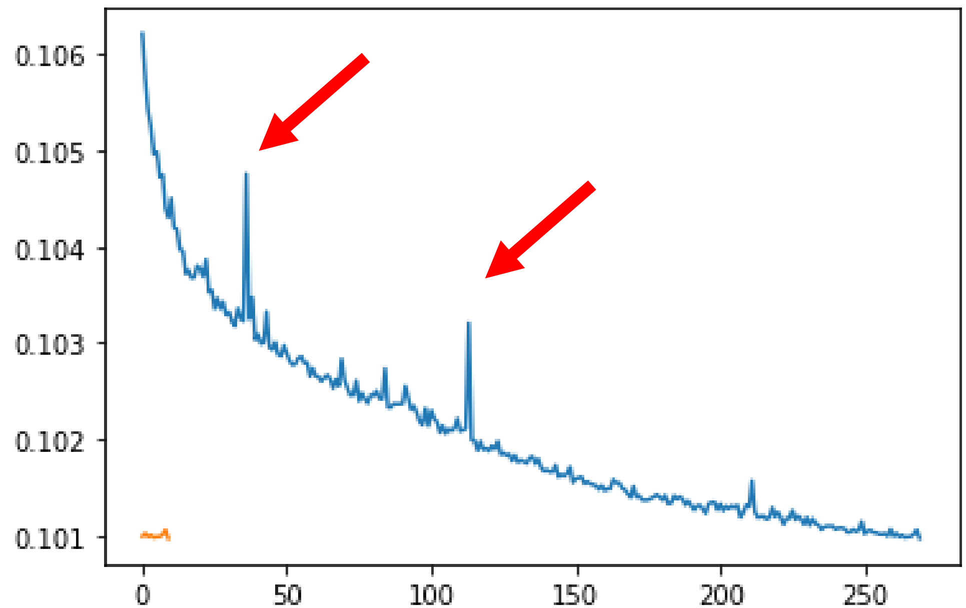
$$q_i = \arg\min_j \left| p_i, r_j \right|_2^2$$

☹

- Normal loss as defined may give a non-convex loss



$chamfer$



$chamfer + normal$

Questions ?

Bibliography

[1] Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W. and Jiang, Y.G., 2018. Pixel2mesh: Generating 3d mesh models from single rgb images. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 52-67).

[2] Dibra, E., Jain, H., Öztireli, C., Ziegler, R. and Gross, M., 2016, October. Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 108-117). IEEE.

[3] Fan, H., Su, H. and Guibas, L.J., 2017. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 605-613).