UC San Diego

Lecture 8:

Deep Learning on Point Cloud

Instructor: Hao Su

Feb 1, 2018

slides credits: Justin Solomon, Chengcheng Tang

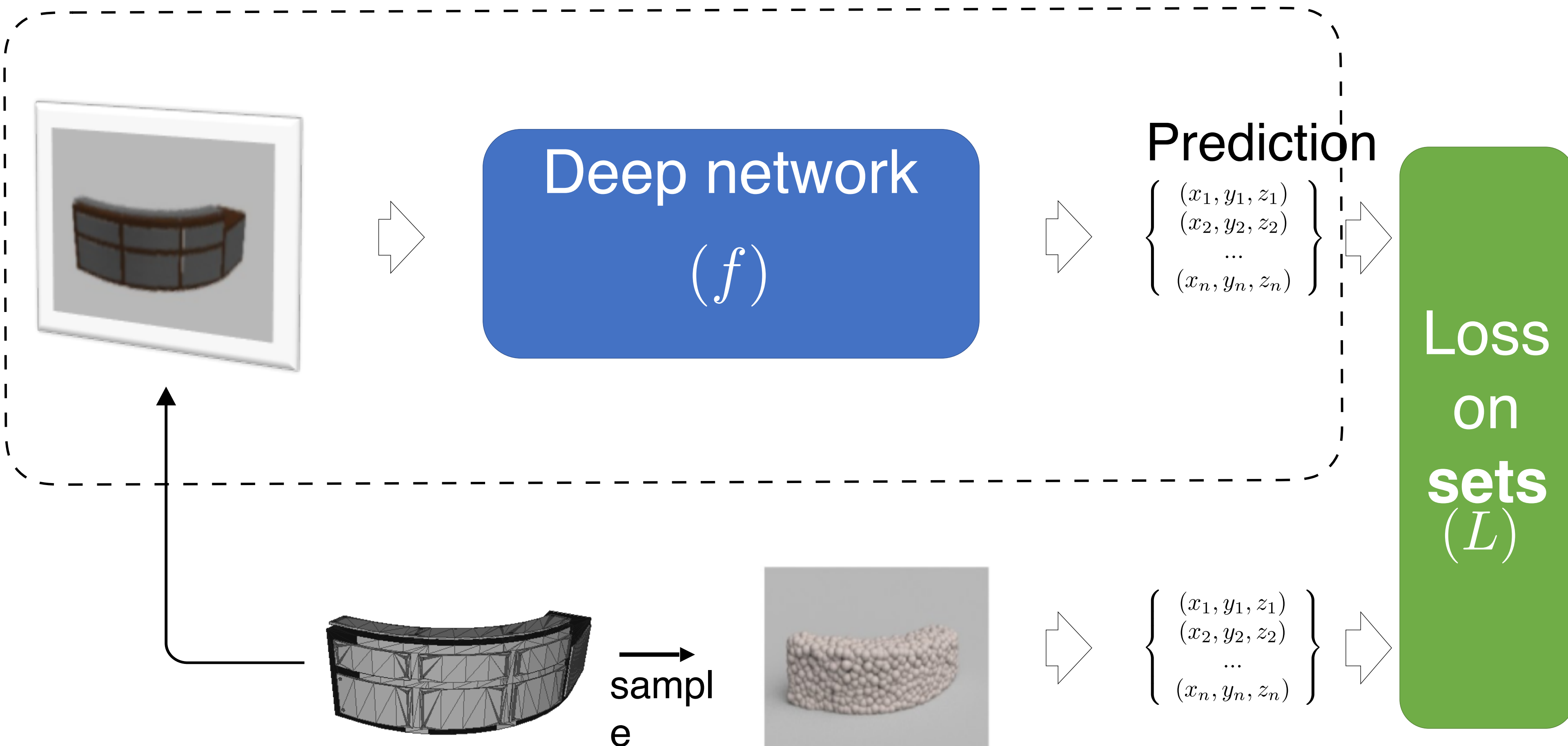
Agenda

- Supervised Point Set Generation (cont)
- Multidimensional Scaling
- Parametric Shape Space for Homotopic Manifolds

Agenda

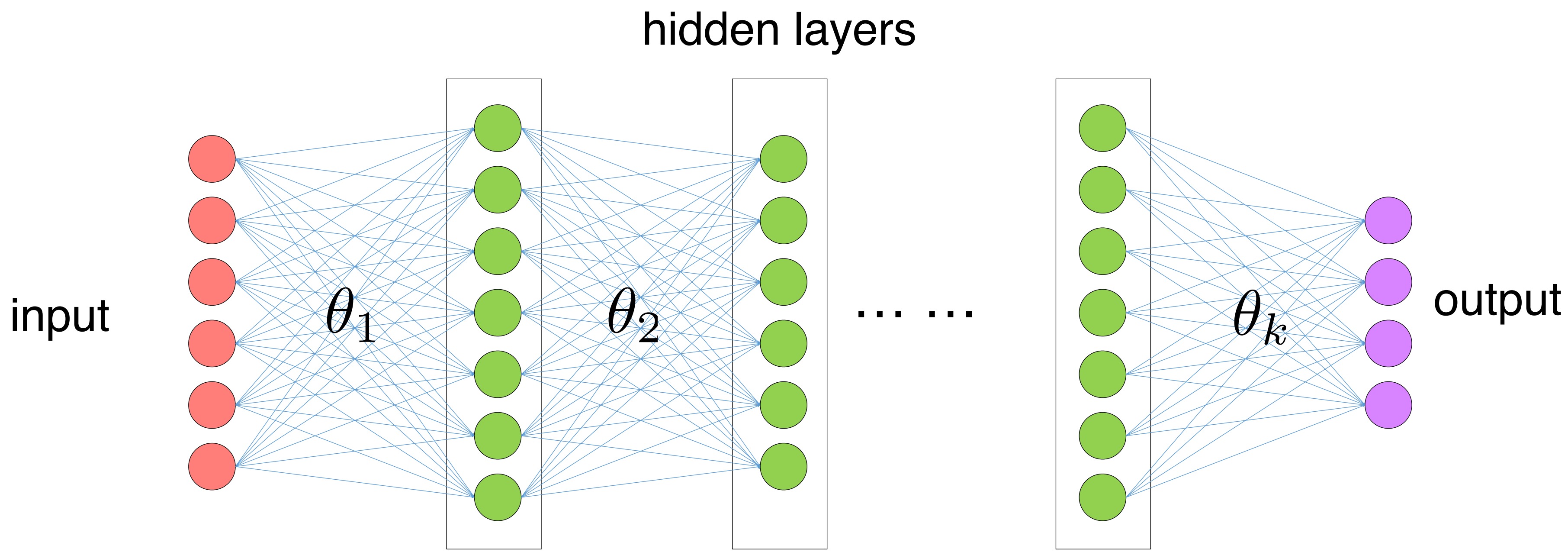
- **Supervised Point Set Generation (cont)**
- Multidimensional Scaling
- Parametric Shape Space for Homotopic Manifolds

Pipeline



CVPR '17, Point Set Generation

Deep neural network

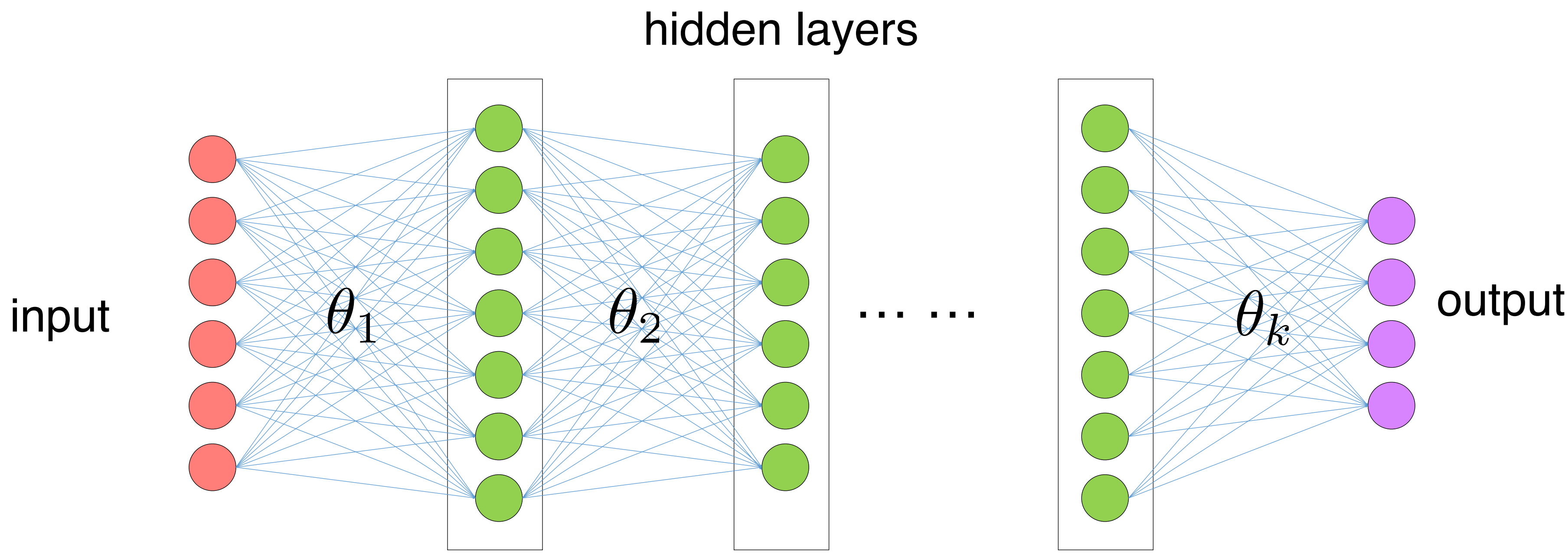


Universal function approximator

- A cascade of layers

CVPR '17, Point Set Generation

Deep neural network

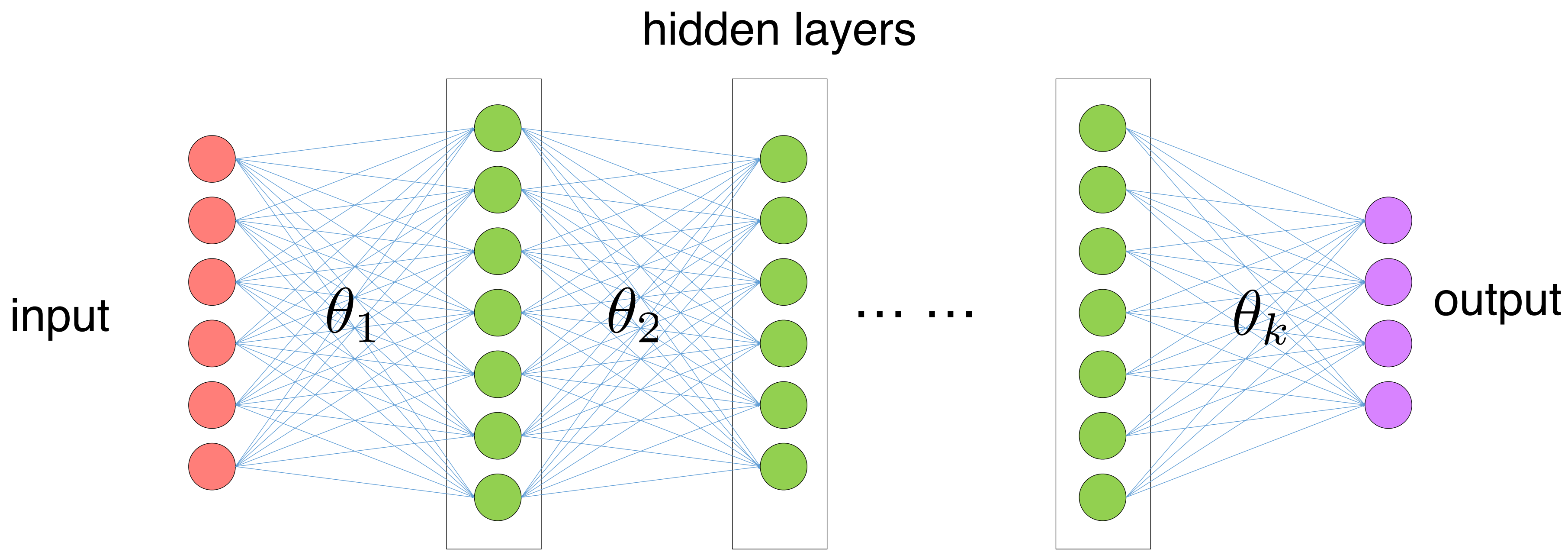


Universal function approximator

- A cascade of layers
- Each layer conducts a simple transformation (parameterized)

CVPR '17, Point Set Generation

Deep neural network

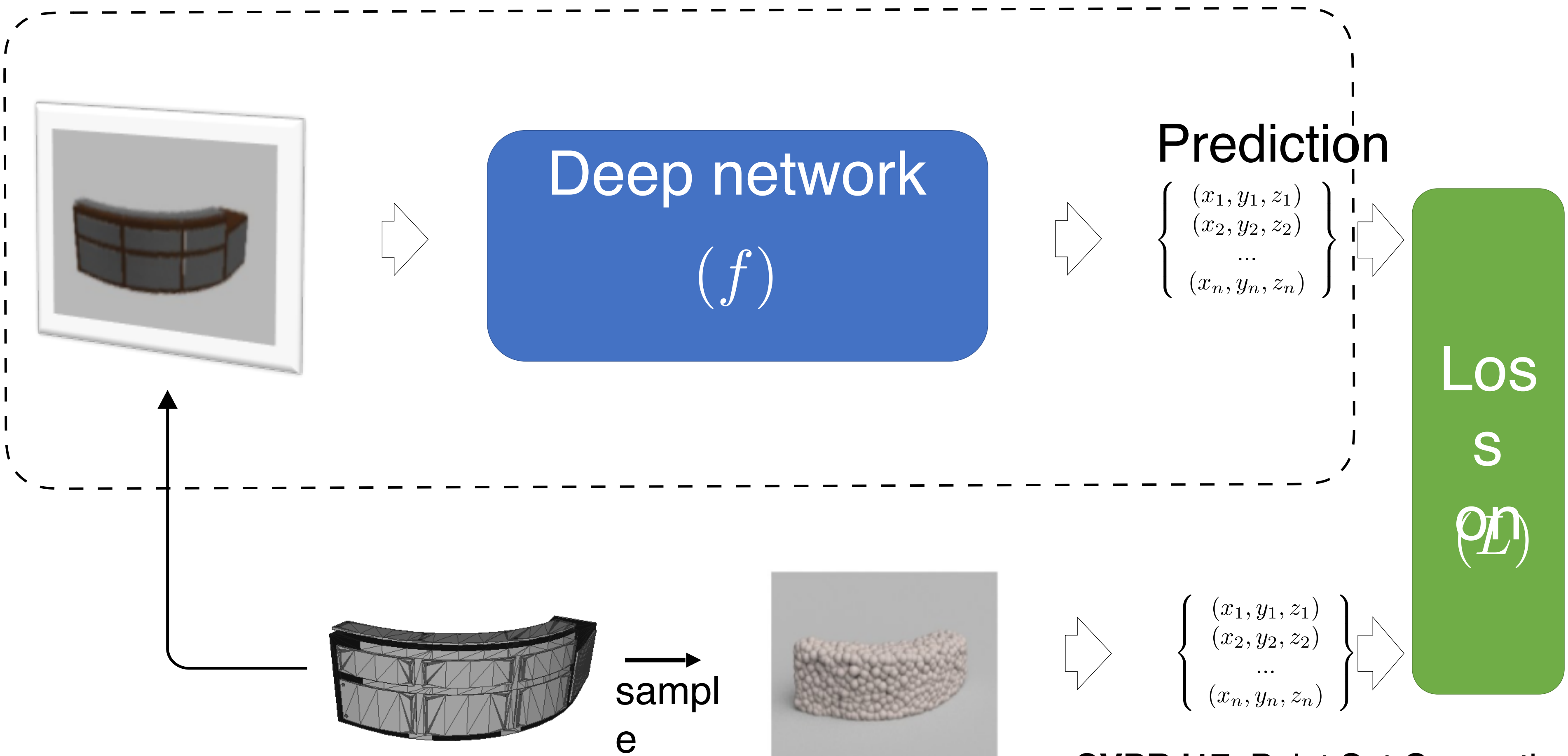


Universal function approximator

- A cascade of layers
- Each layer conducts a simple transformation (parameterized)
- Millions of parameters, has to be fitted by many data

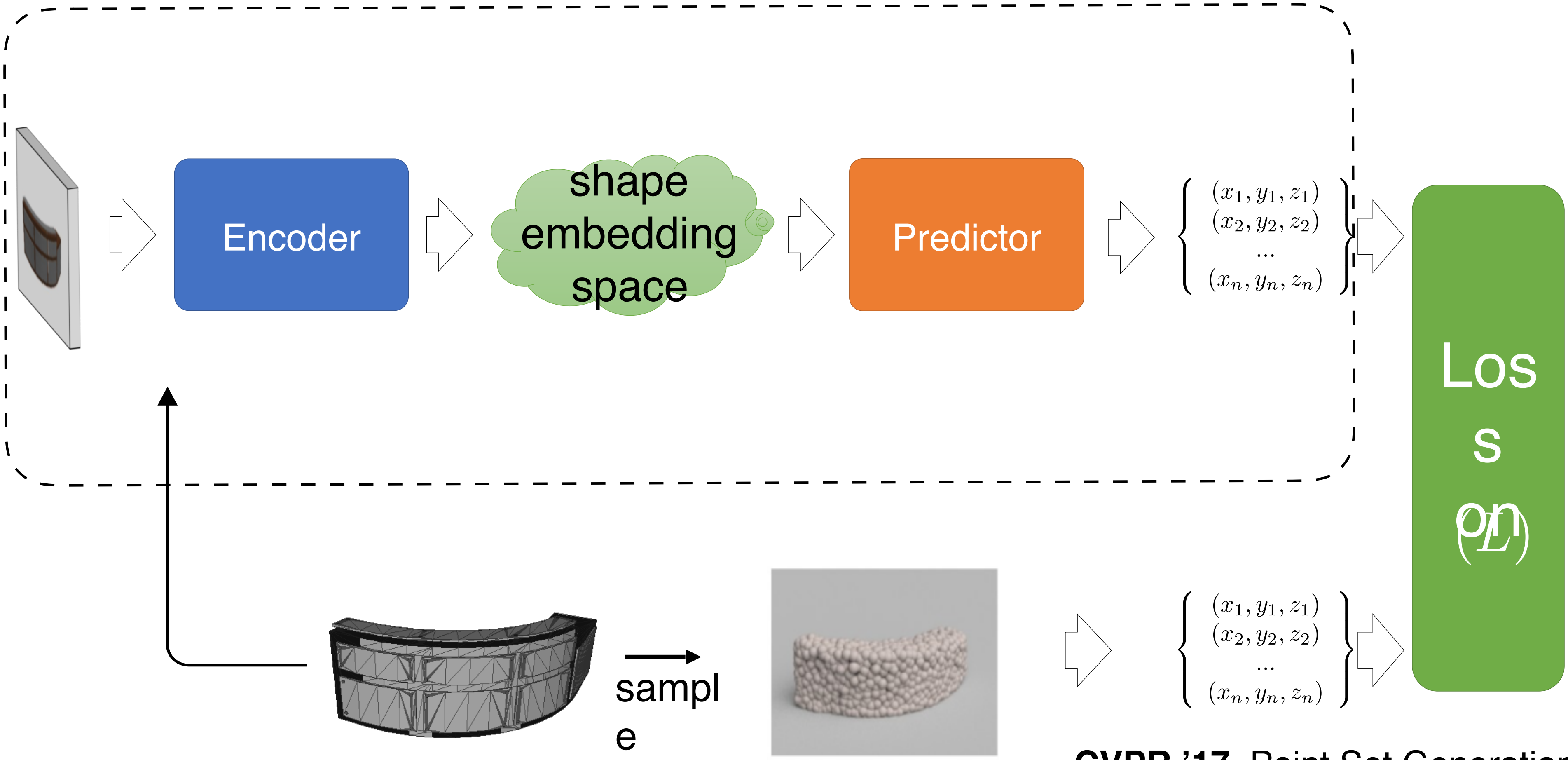
CVPR '17, Point Set Generation

Pipeline



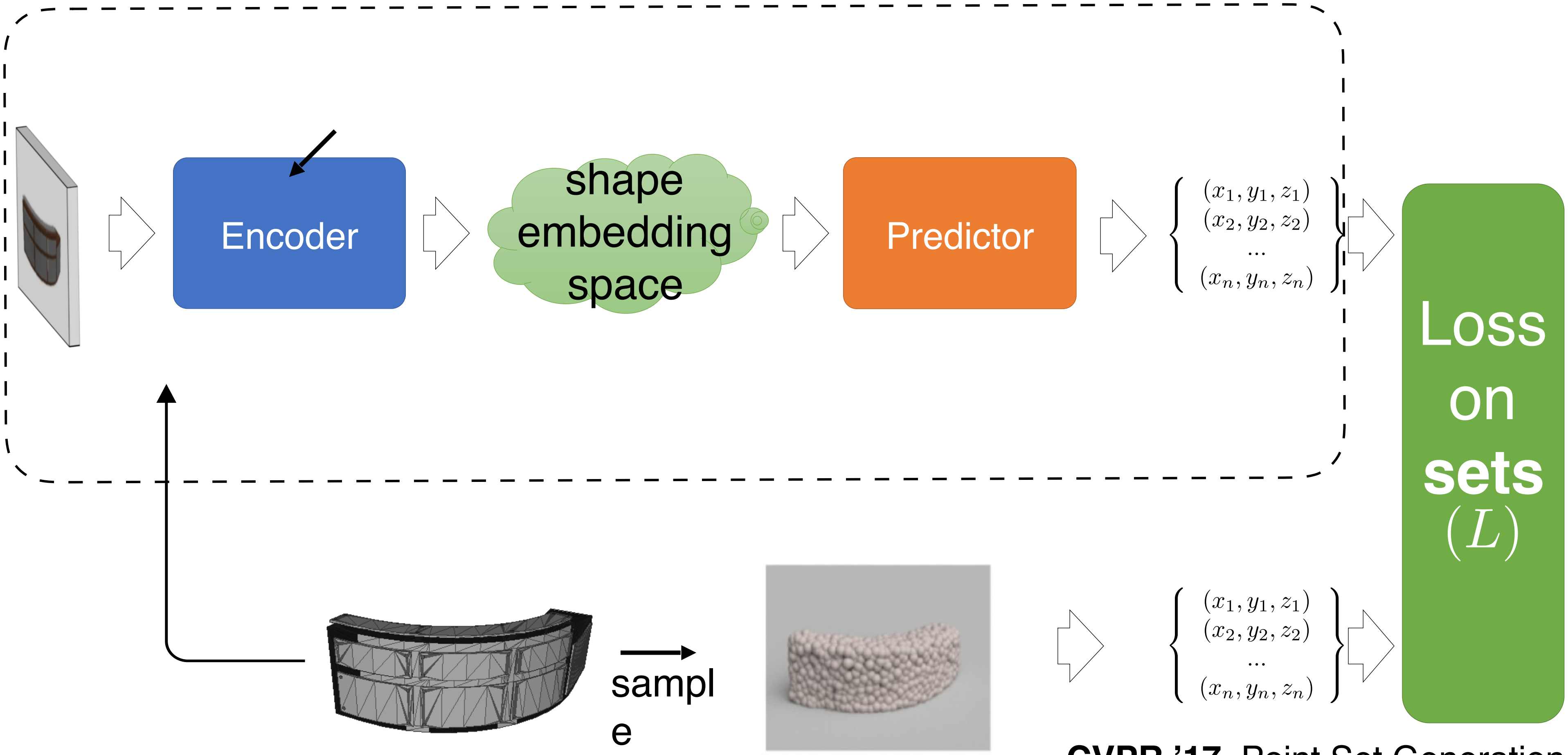
CVPR '17, Point Set Generation

Pipeline



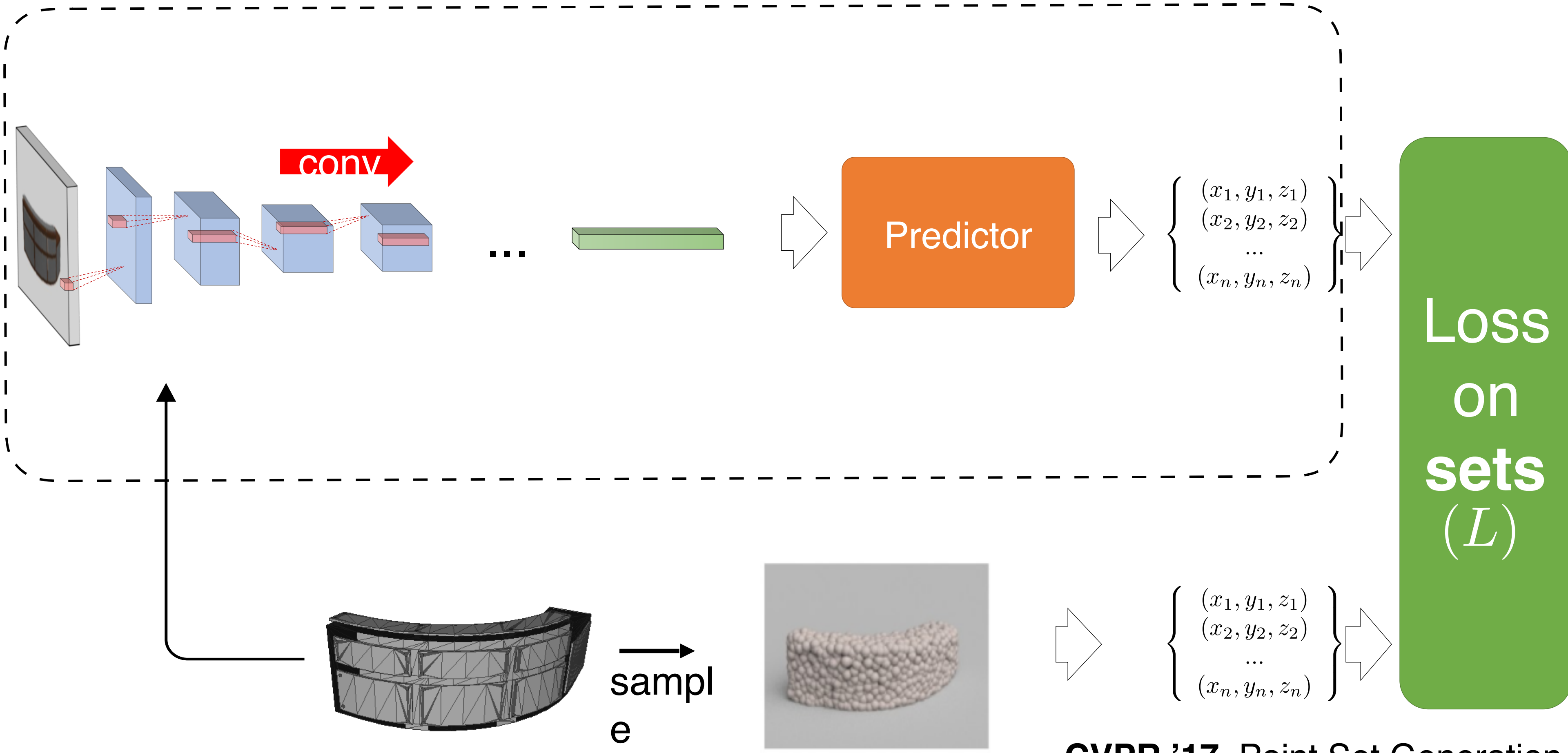
CVPR '17, Point Set Generation

Pipeline



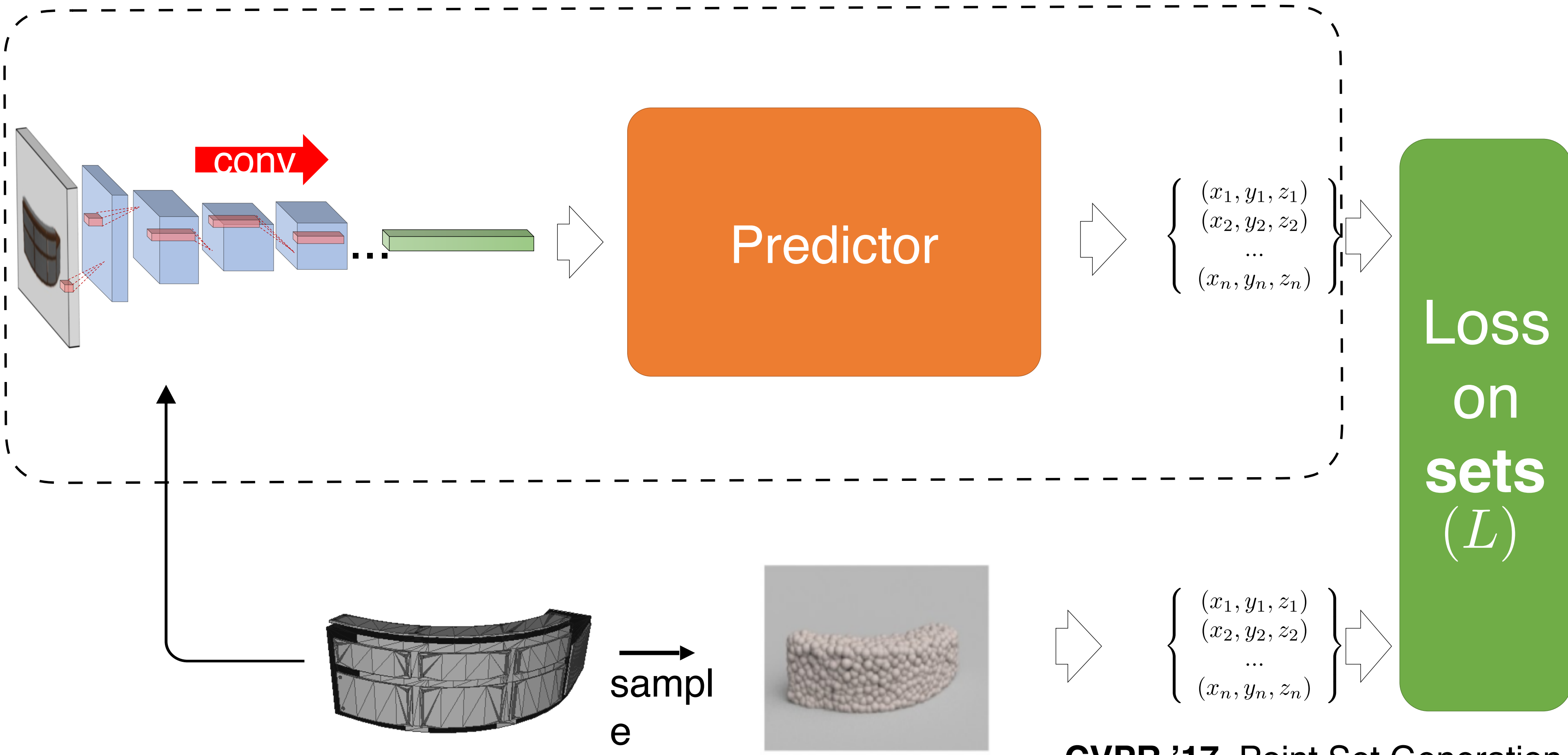
CVPR '17, Point Set Generation

Pipeline



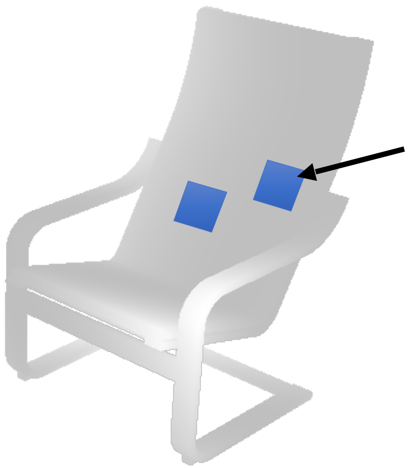
CVPR '17, Point Set Generation

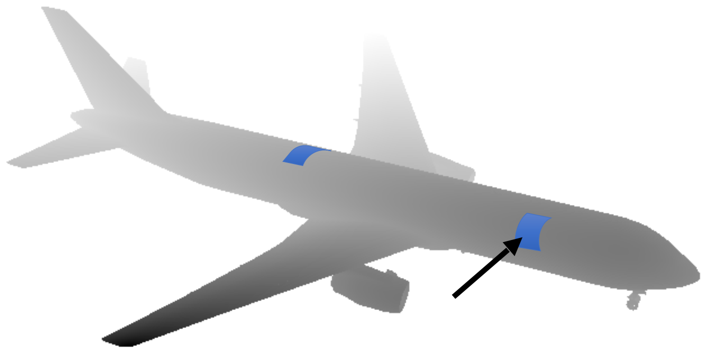
Pipeline



CVPR '17, Point Set Generation

Natural statistics of geometry

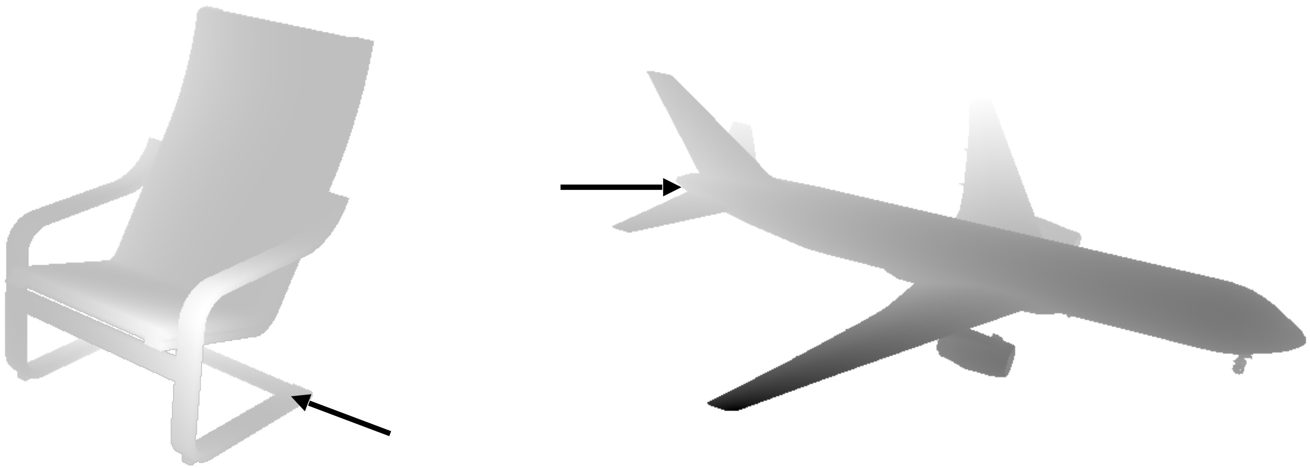




- Many local structures are common
  - e.g., planar patches, cylindrical patches
  - **strong local correlation** among point coordinates

CVPR '17, Point Set Generation

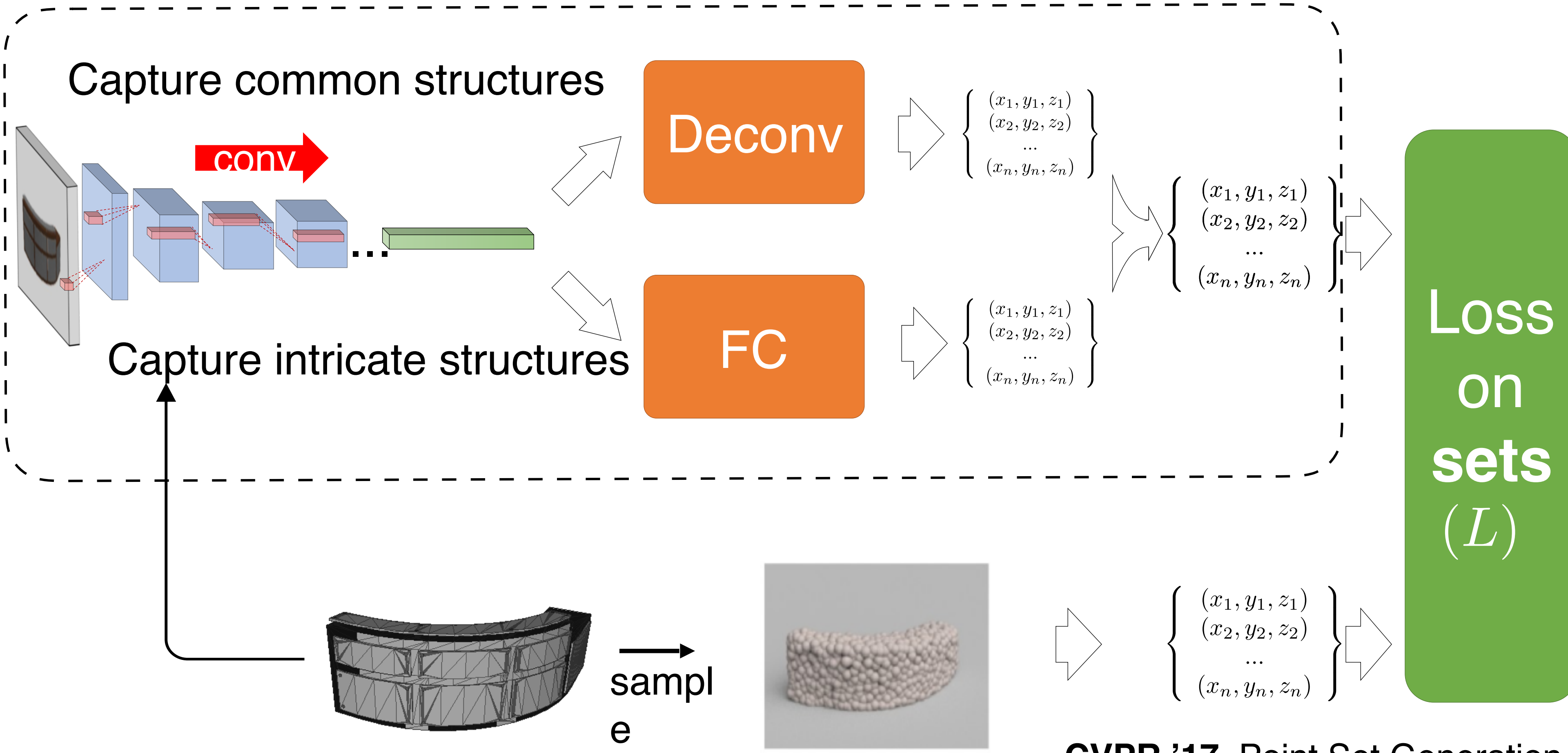
Natural statistics of geometry



- Many local structures are common
  - e.g., planar patches, cylindrical patches
  - **strong local correlation** among point coordinates
- Also some intricate structures
  - points have **high local variation**

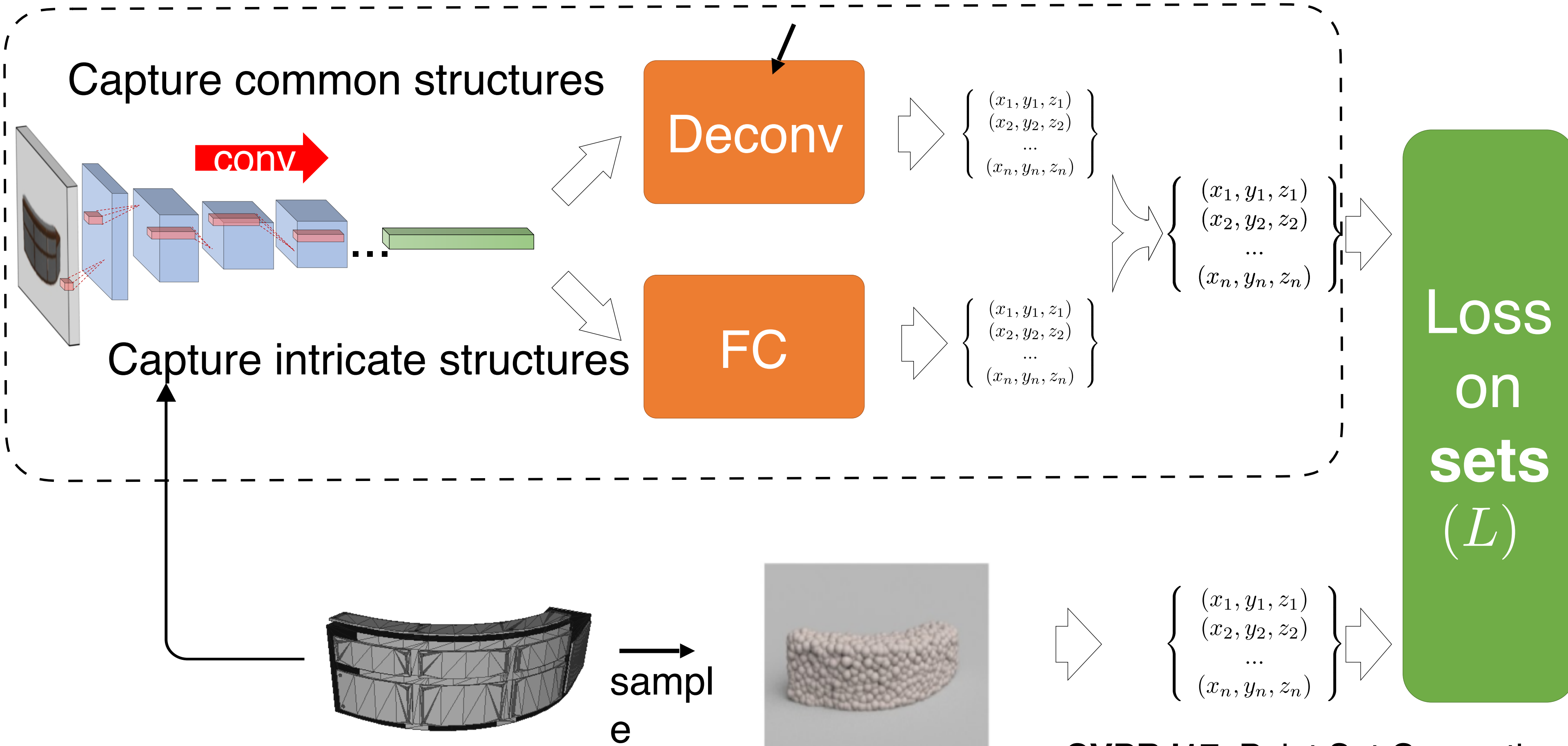
CVPR ’17, Point Set Generation

Pipeline



CVPR '17, Point Set Generation

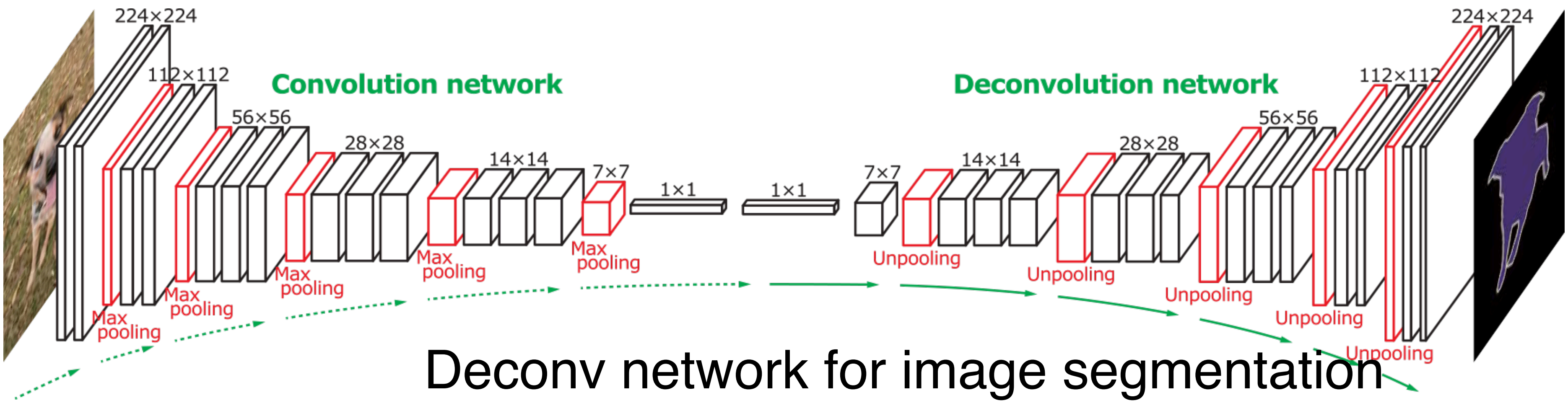
Pipeline



CVPR '17, Point Set Generation

Review: deconv network

- Output D arrays, e.g., 2D segmentation map
- **Common local patterns** are **learned from data**
- Predict **locally correlated** data well
- Weight sharing reduces the number of params

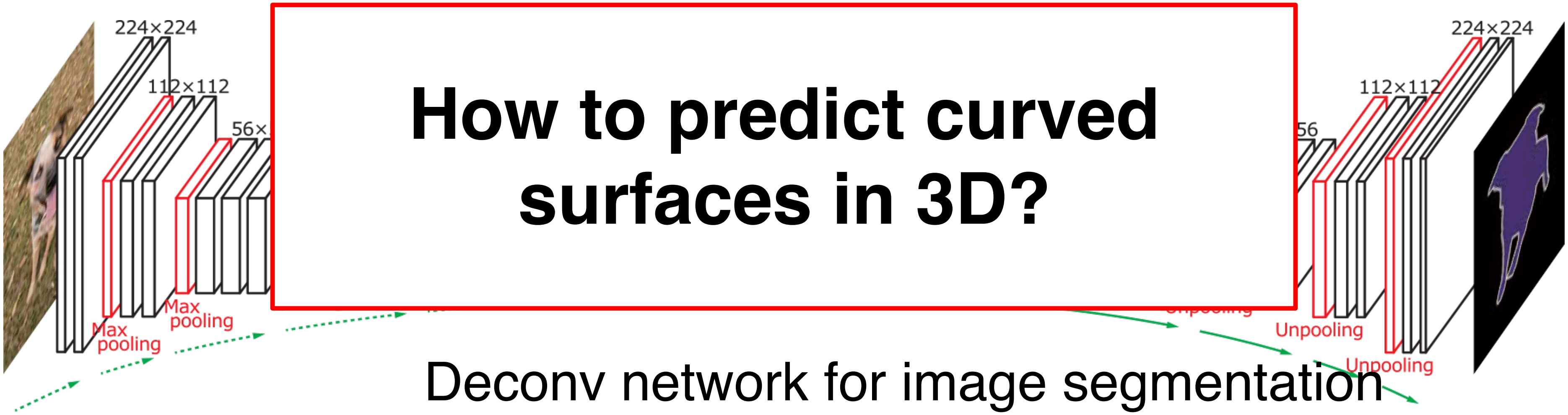


Deconv network for image segmentation

Credit: FCNN, Long

Review: deconv network

- Output n D arrays, e.g., 2D segmentation map
- **Common local patterns** are **learned from data**
- Predict **locally correlated** data well
- Weight sharing reduces the number of params



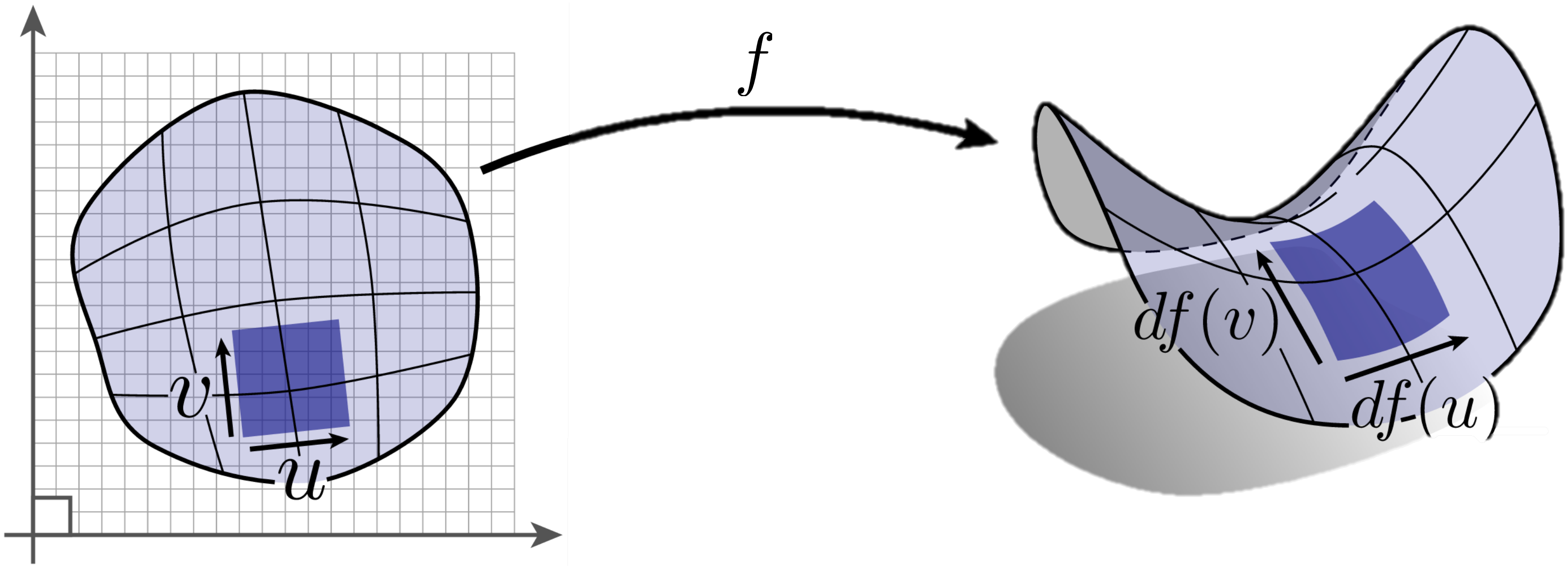
How to predict curved surfaces in 3D?

Deconv network for image segmentation

Credit: FCNN, Long

Prediction of curved 2D surfaces in 3D

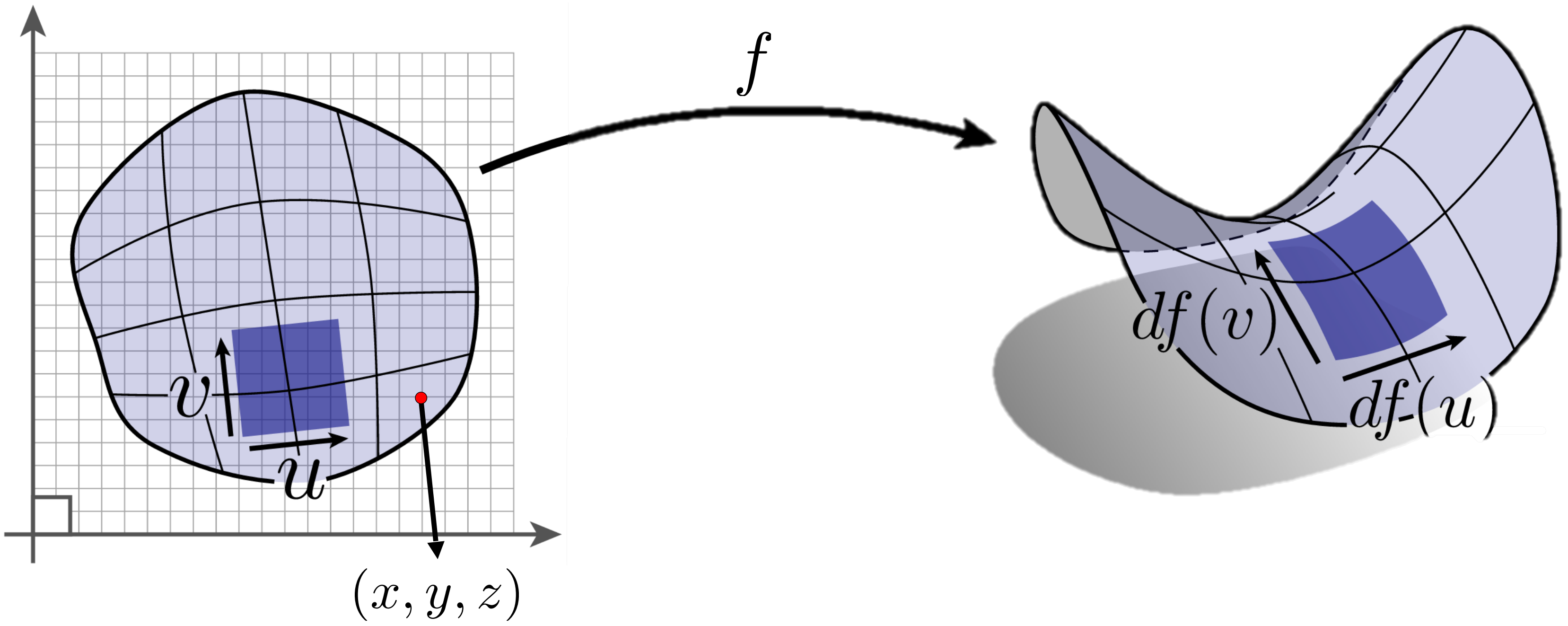
- Surface parametrization (2D $\leftrightarrow$ 3D mapping)



Credit: Discrete Differential Geometry,

Prediction of curved 2D surfaces in 3D

- Surface parametrization (2D $\leftrightarrow$ 3D mapping)

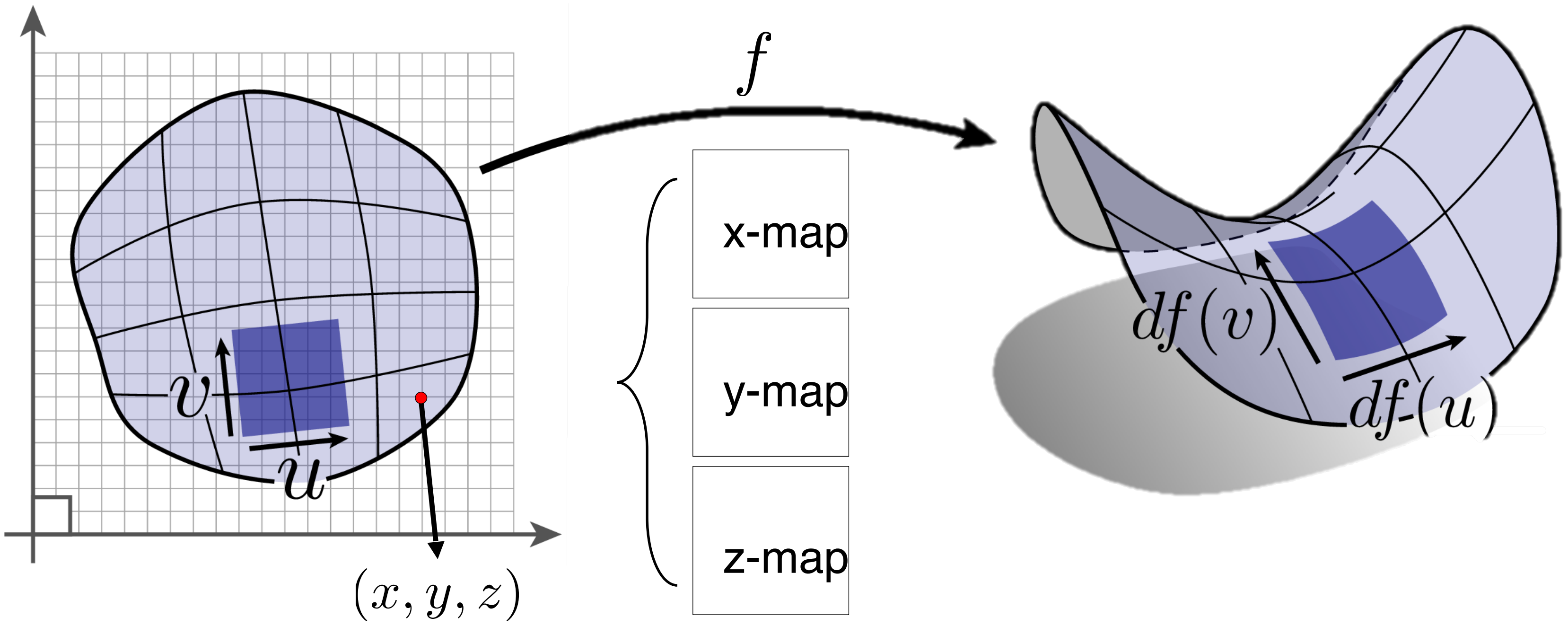


Credit: Discrete Differential Geometry,

Prediction of curved 2D surfaces in 3D

- Surface parametrization (2D-3D mapping)

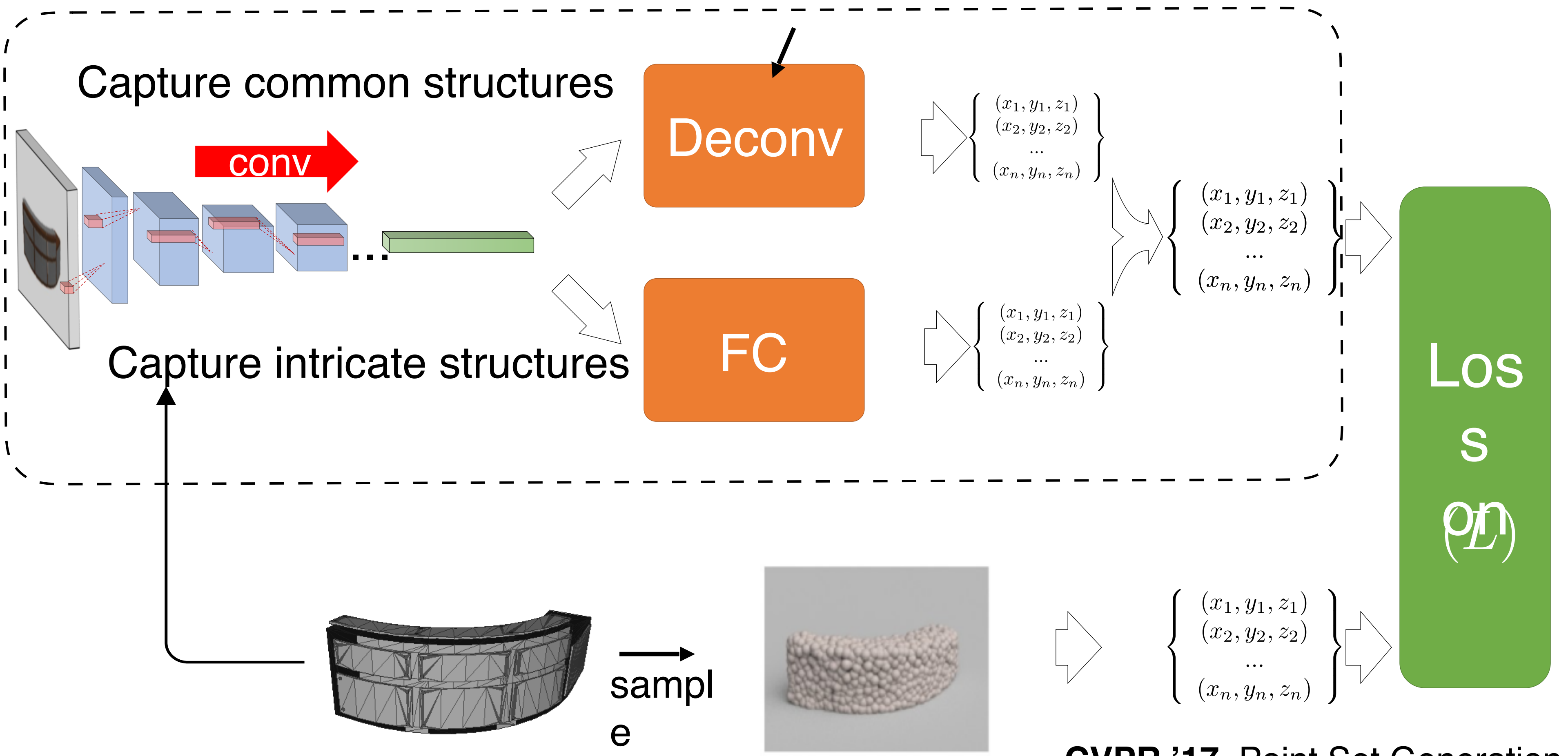
$\leftrightarrow$



coordinate maps

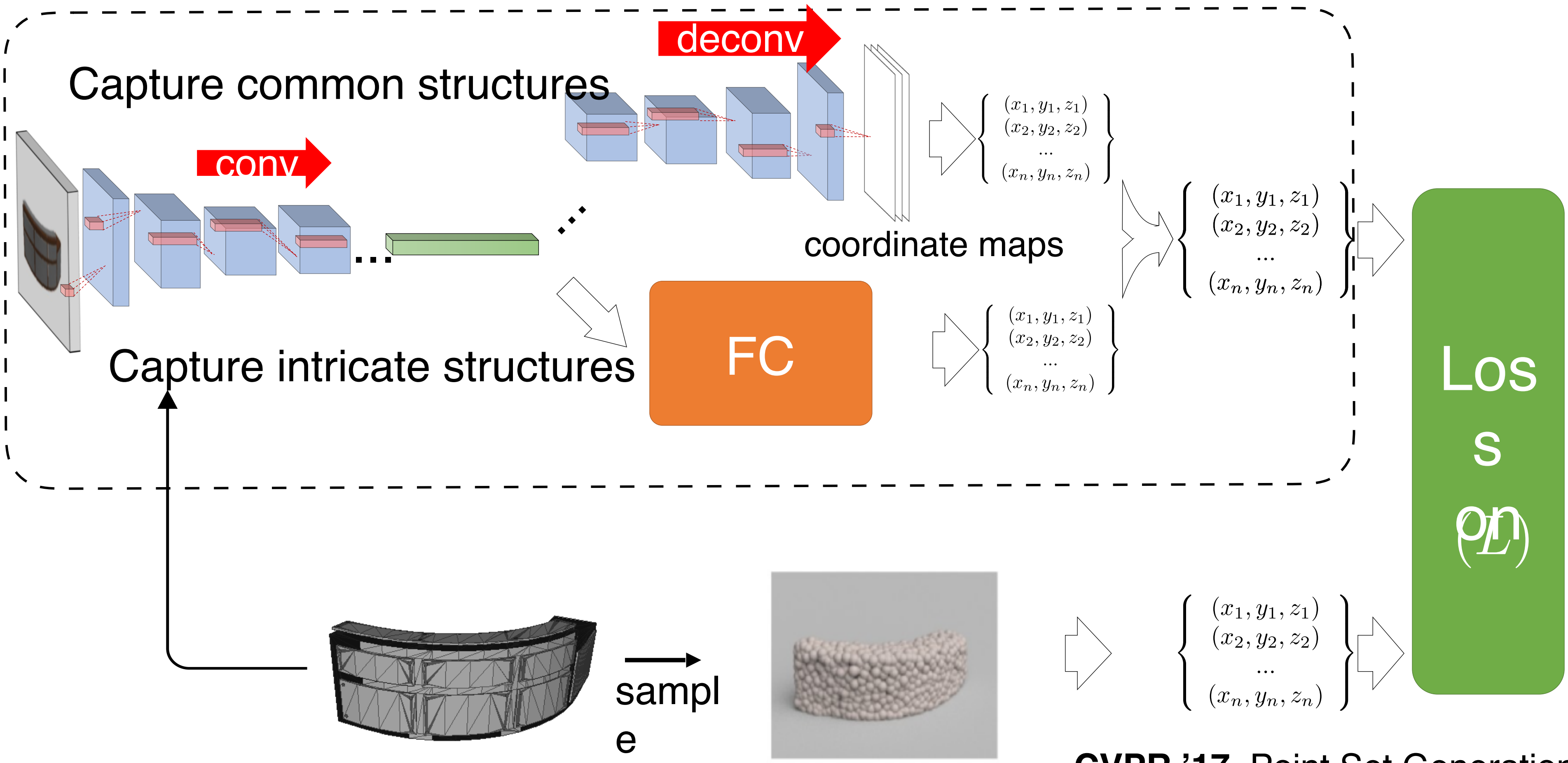
Credit: Discrete Differential Geometry, Crane et al.

Parametrization prediction by deconv network



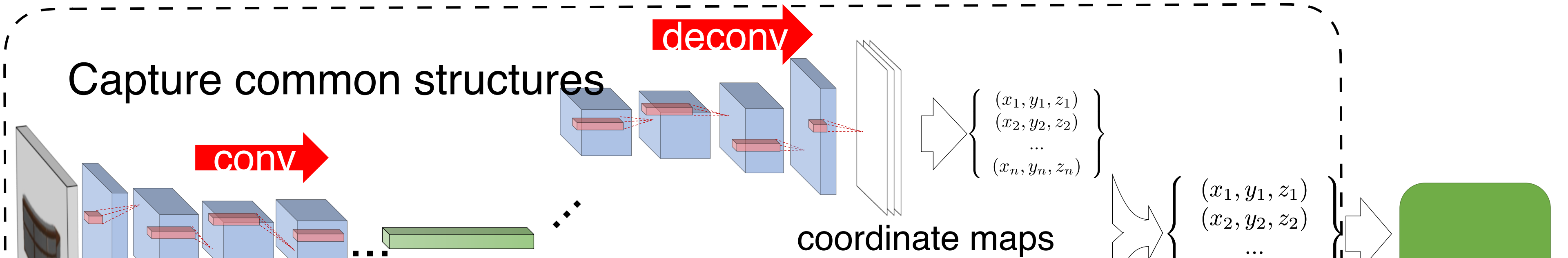
CVPR '17, Point Set Generation

Parametrization prediction by deconv network



CVPR '17, Point Set Generation

Parametrization prediction by deconv network



Note that

- The parametrization (2D/3D mapping) is learned from data
- i.e., obtains a network and data friendly parametrization

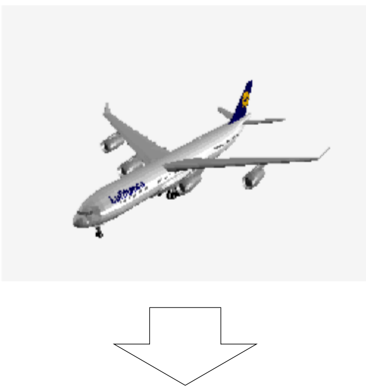
CVPR '17, Point Set Generation

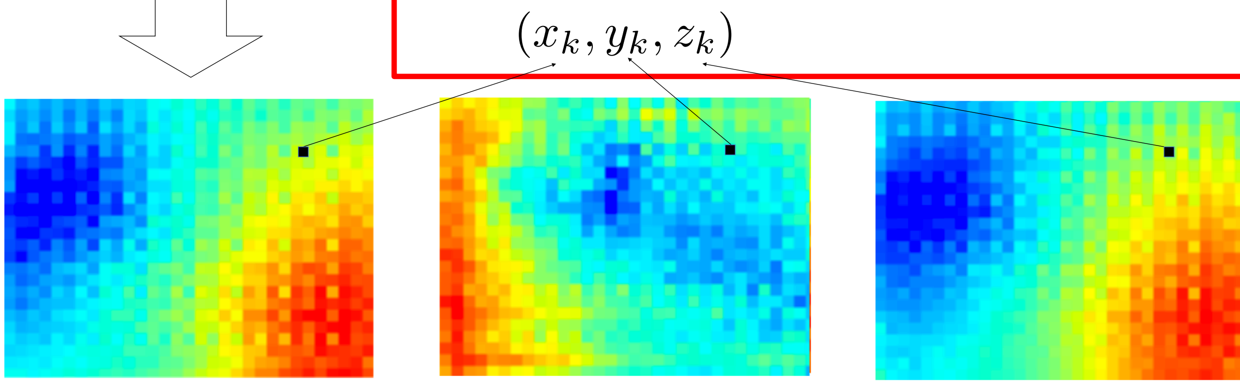
Visualization of the learned parameterization

- Surface parametrization (2D   3D mapping)

Observation:

- Learns a **smooth** parametrization
- Because deconv net tends to predict data with local correlation

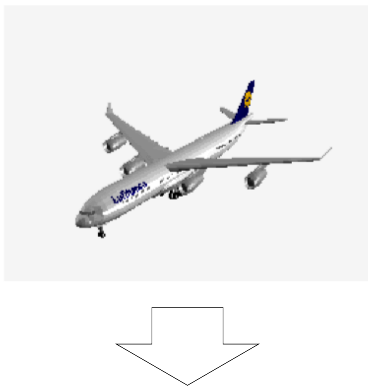




map of x coord map of y coord map of z coord

Visualization of the learned parameterization

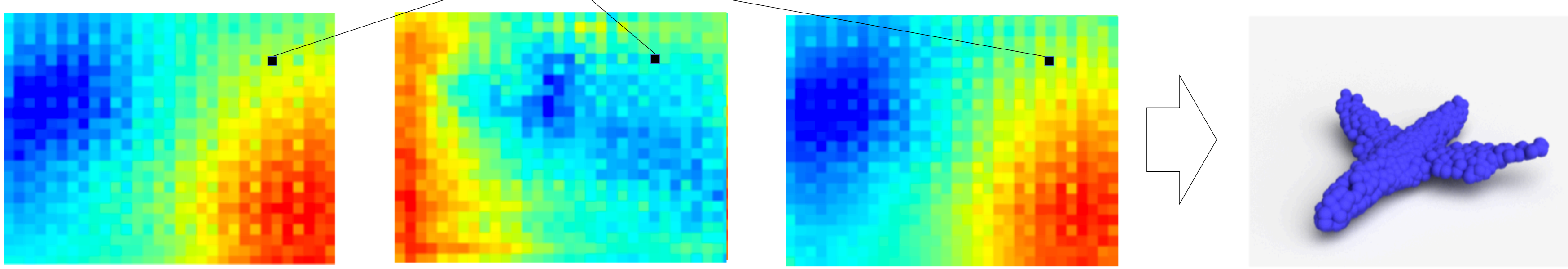
- Surface parametrization (2D 3D mapping)



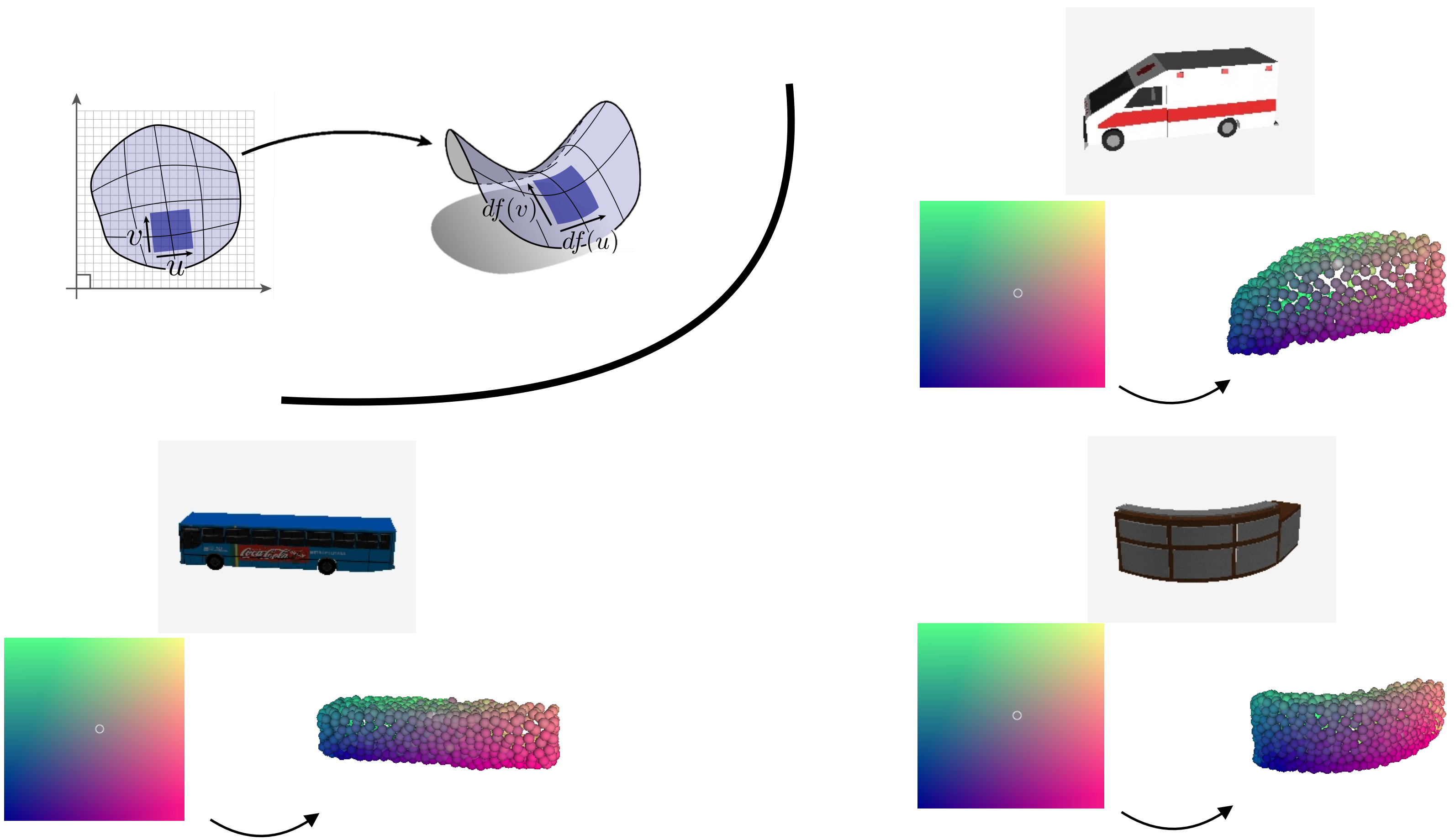
Observation:

- Learns a **smooth** parametrization
- Because deconv net tends to predict data with local correlation

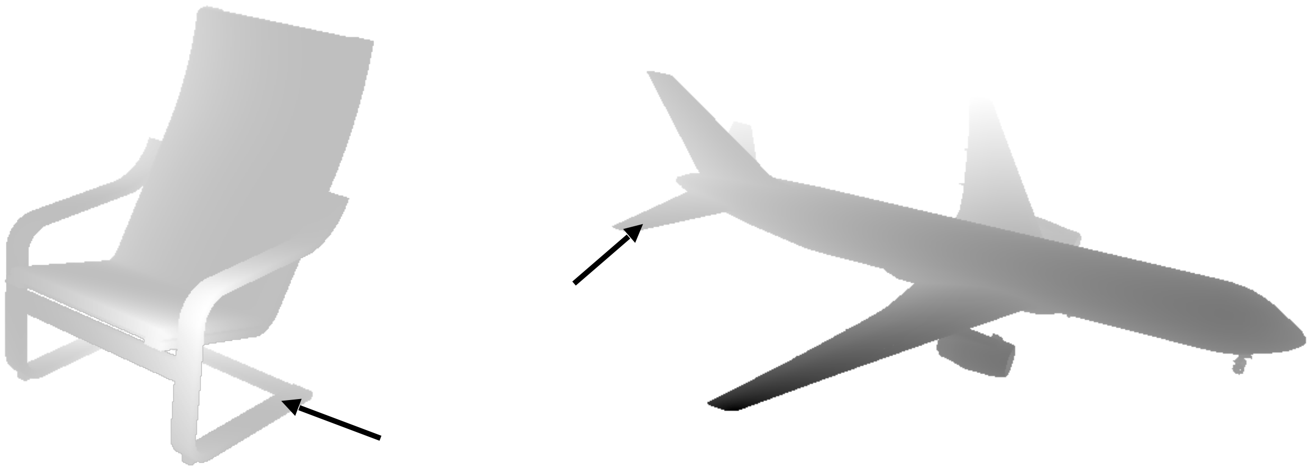
(x_k, y_k, z_k)



map of x coord map of y coord map of z coord


v
u
df(v)
df(u)

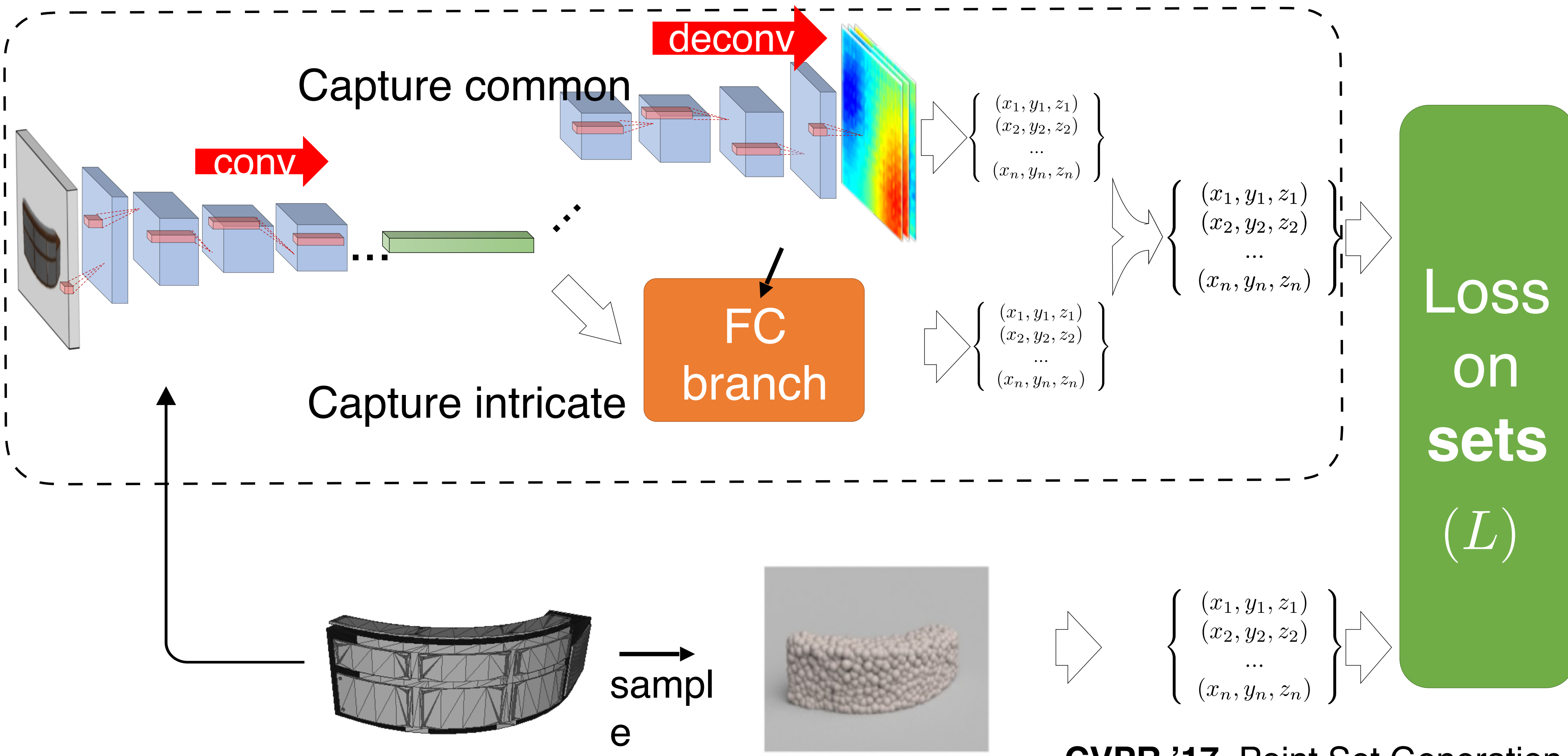
Natural statistics of geometry



- Many local structures are common
  - e.g., planar patches, cylindrical patches
  - **strong local correlation** among point coordinates
- Also some intricate structures

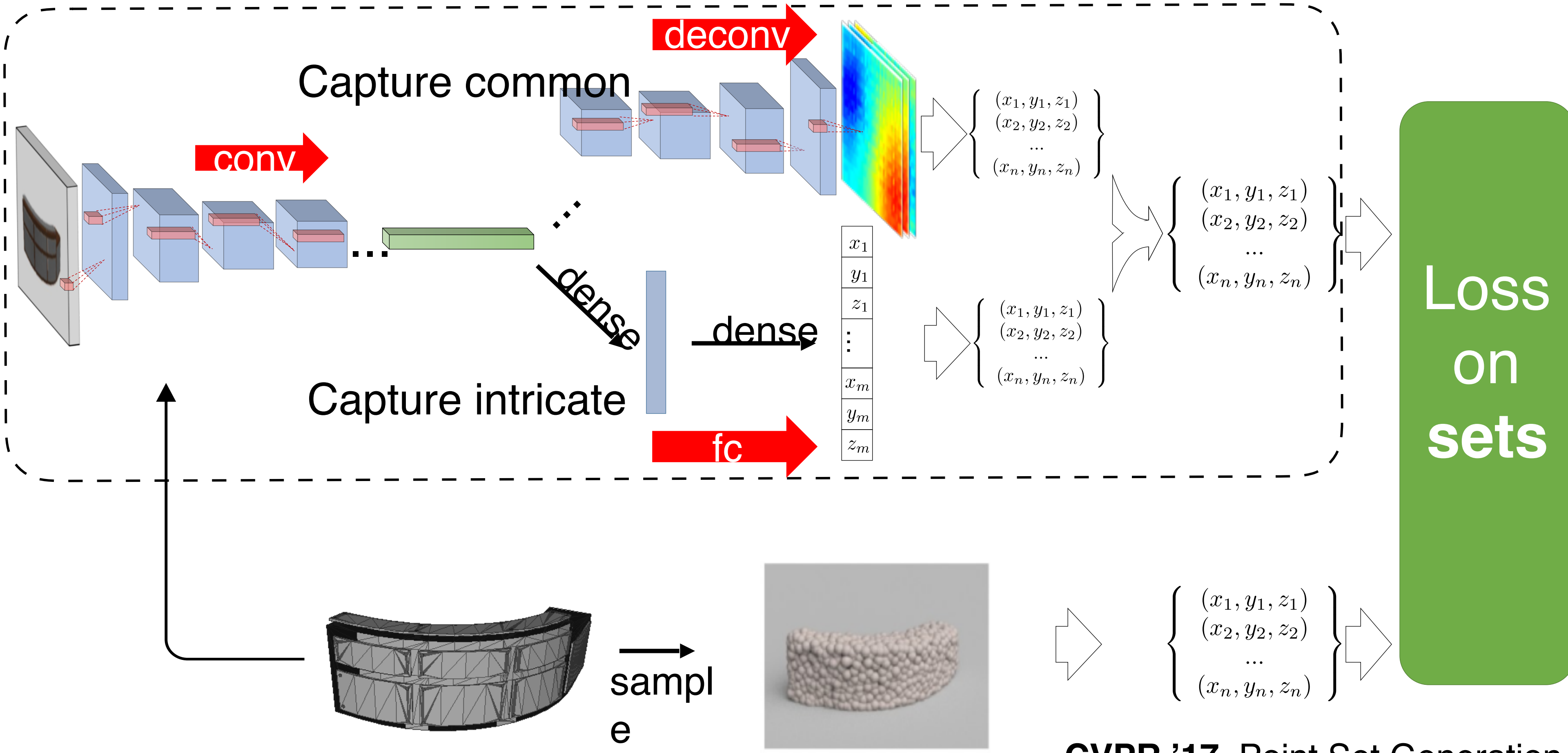
CVPR '17, Point Set Generation

Pipeline



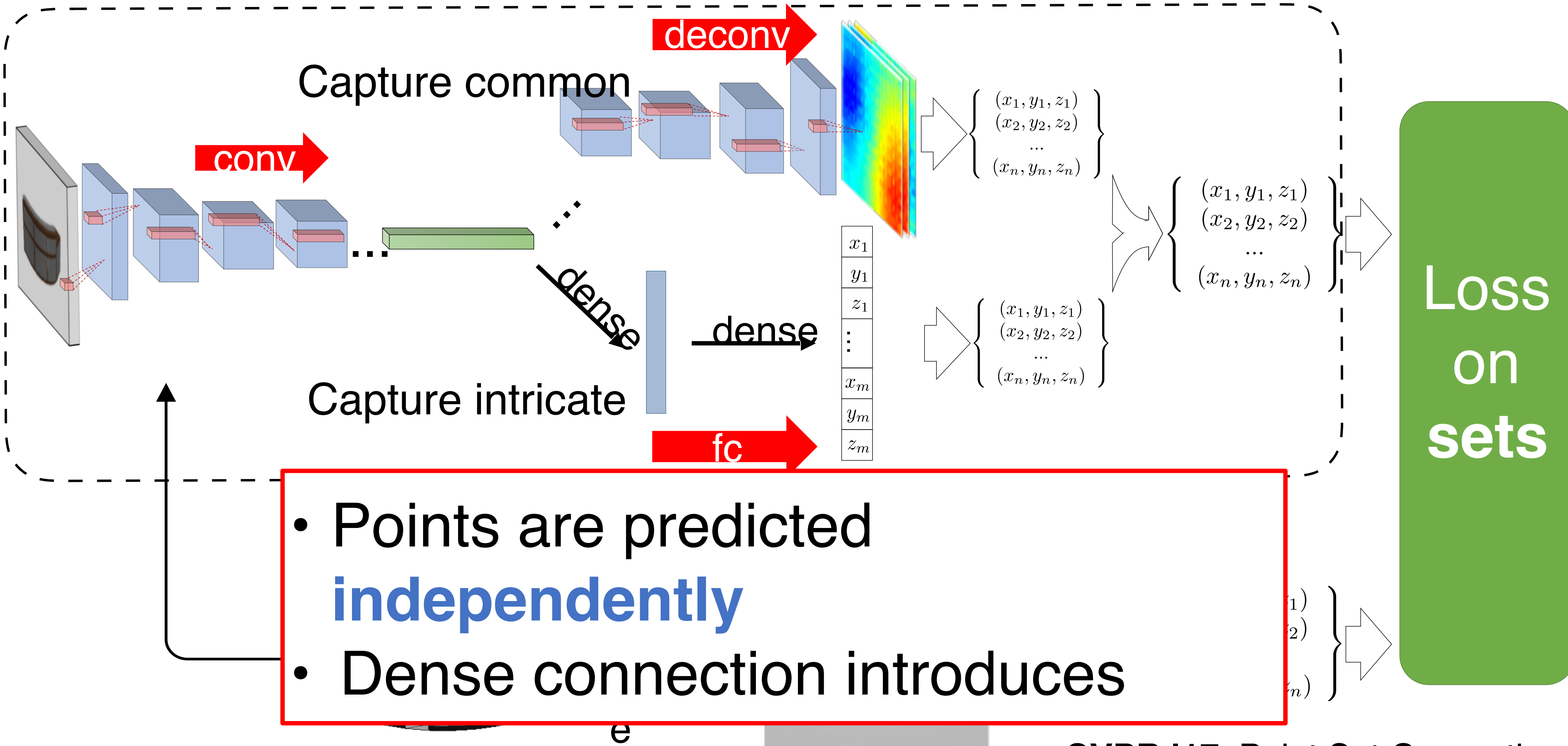
CVPR '17, Point Set Generation

Pipeline



CVPR ’17, Point Set Generation

Pipeline



- Points are predicted **independently**
- Dense connection introduces

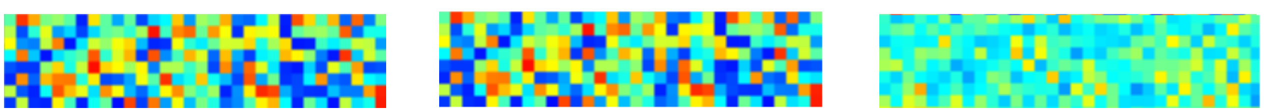
CVPR '17, Point Set Generation

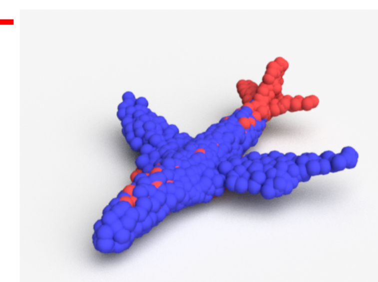
Visualization of the effect of FC branch

- Surface parametrization (2D 3D mapping)

Observation:

- The arrangement of predicted points are uncorrelated



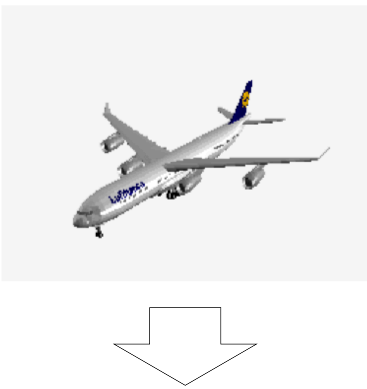


x-coord y-coord z-coord **red**

CVPR '17, Point Set Generation

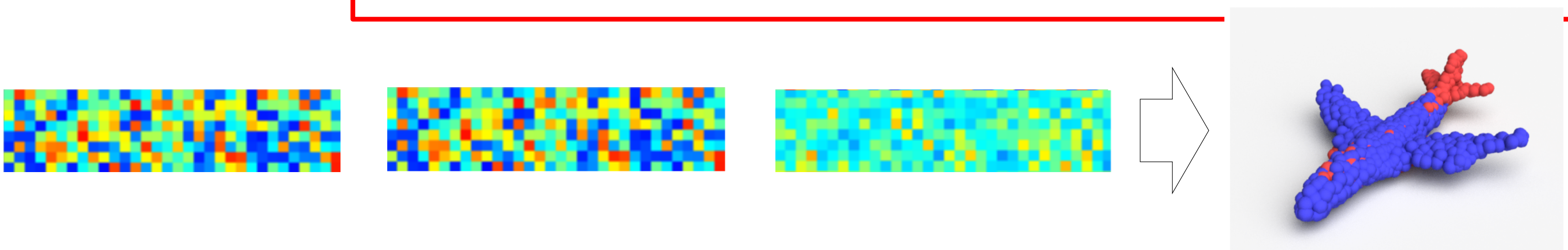
Visualization of the effect of FC branch

- Surface parametrization (2D   3D mapping)



Observation:

- The arrangement of predicted points are uncorrelated
- Located at **fine** structures



CVPR ’17, Point Set Generation

Q: Which color corresponds to the deconv branch? FC branch?

CVPR ’17, Point Set Generation

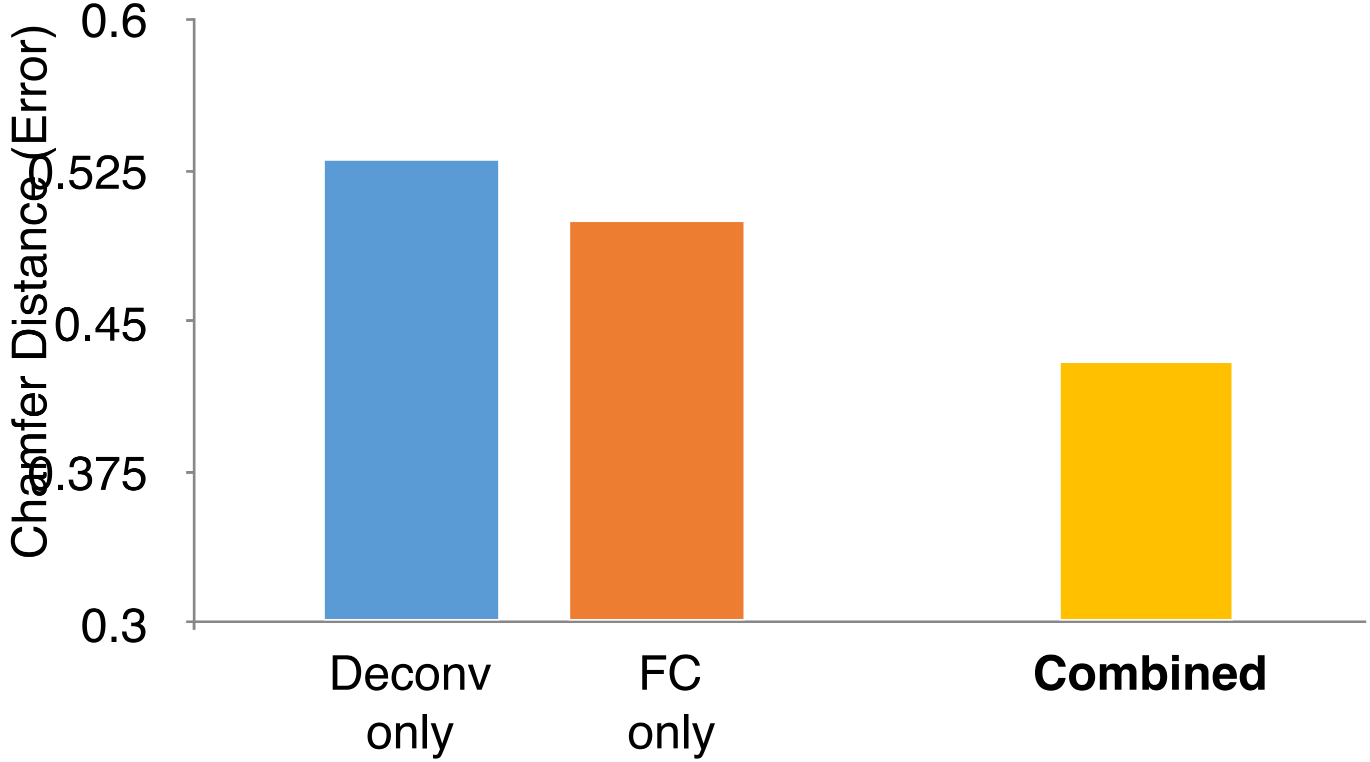
Q: Which color corresponds to the deconv branch? FC branch?

blue: deconv branch – **large, smooth** structures
red: FC branch – **intricate** structures

CVPR '17, Point Set Generation

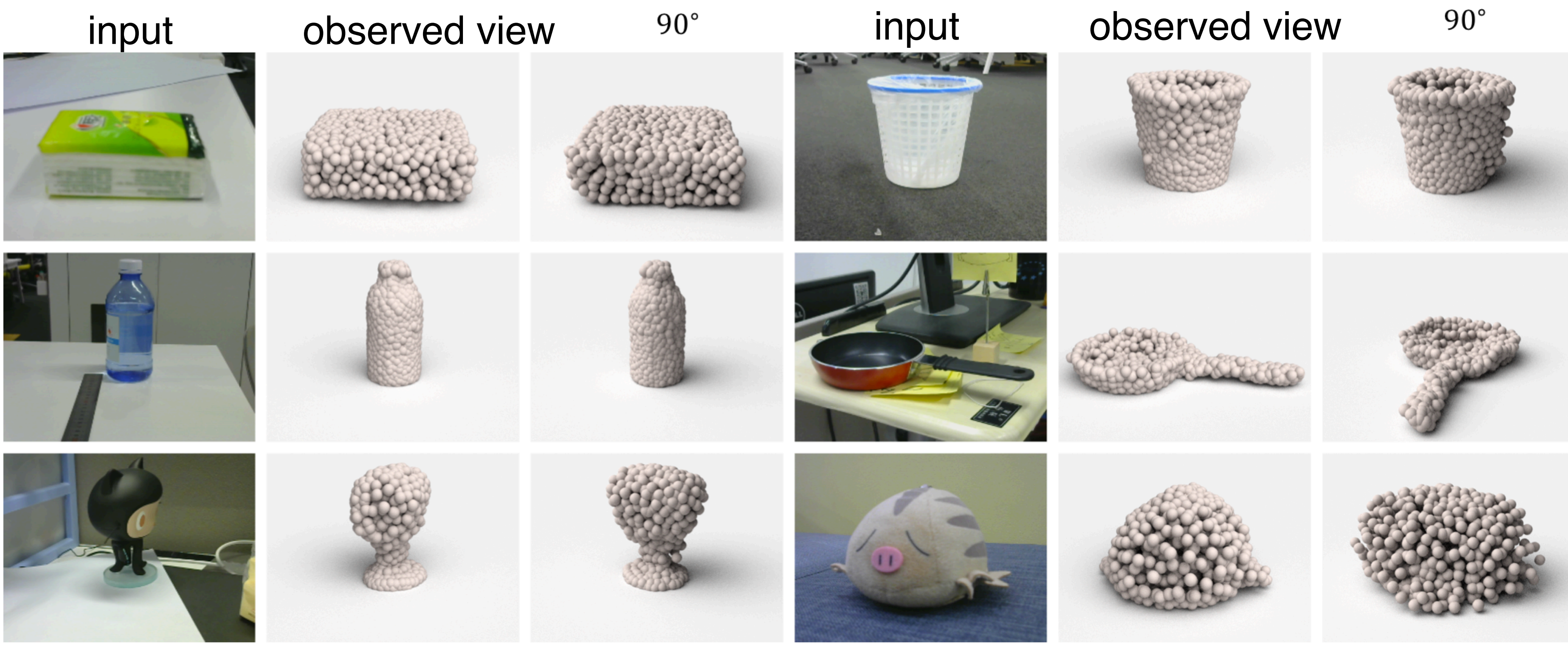
Effect of combining two branches

Train/tested on 2K object categories



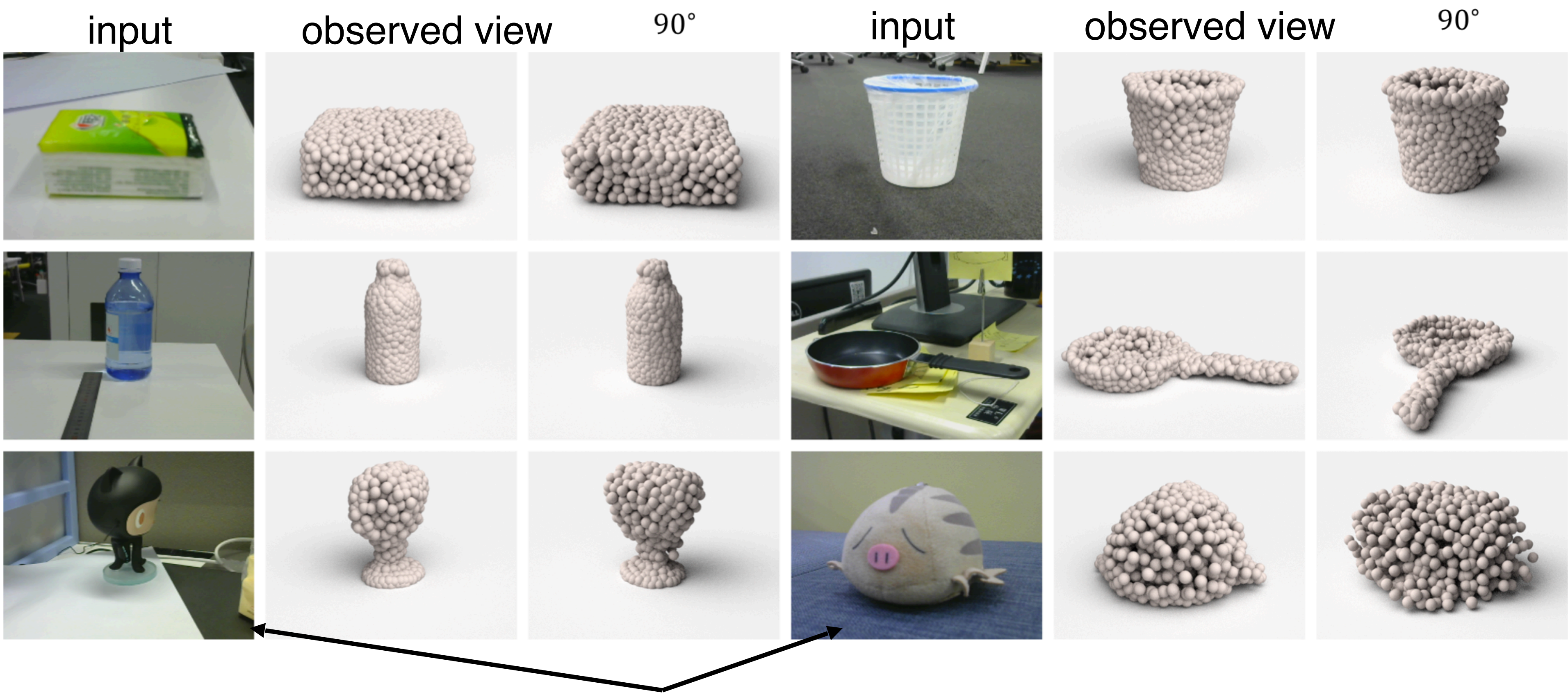
CVPR '17, Point Set Generation

Real-world results



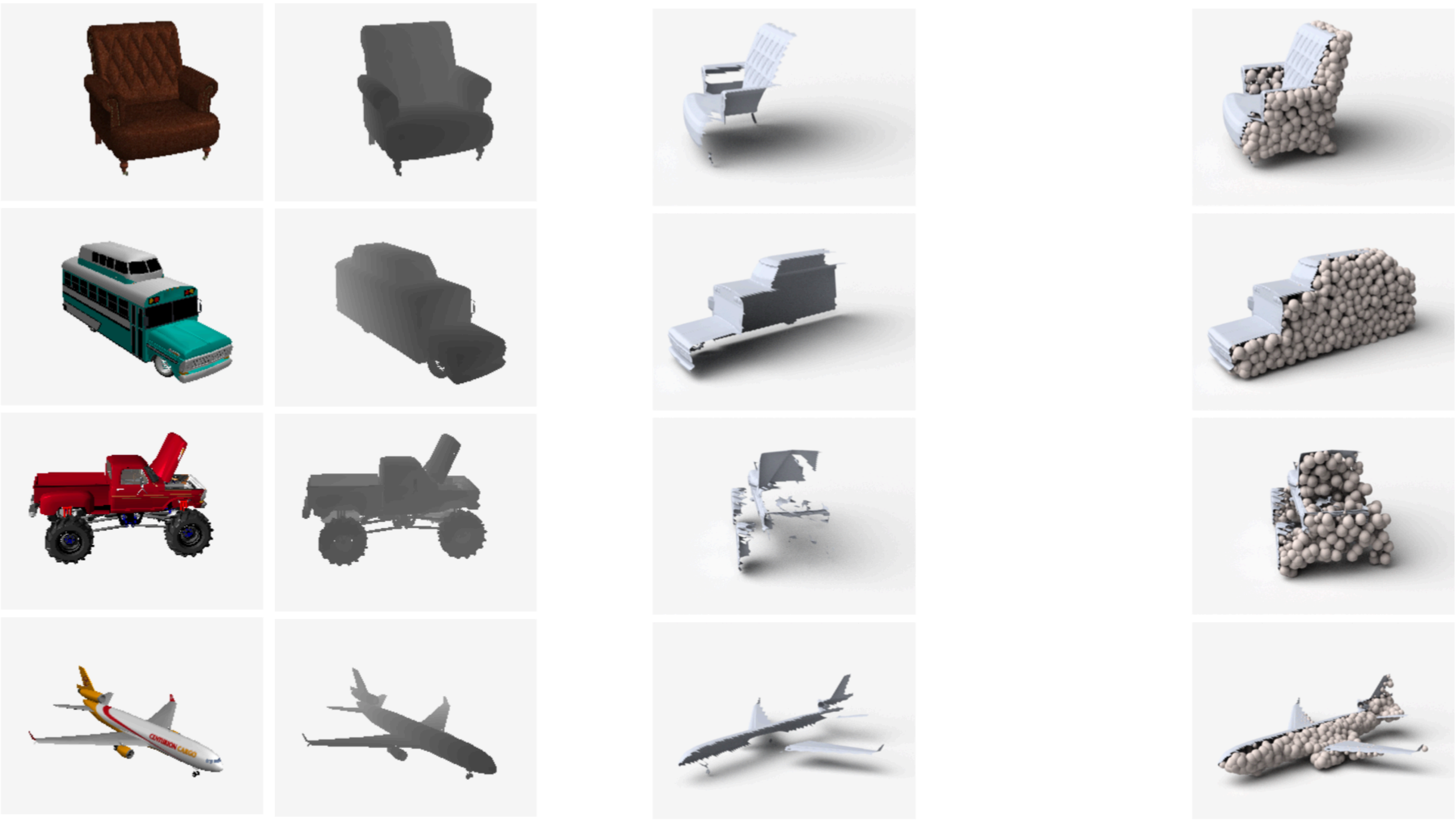
CVPR '17, Point Set Generation

Generalization to unseen categories



CVPR '17, Point Set Generation

Extension: shape completion for RGBD data



RGBD map (input) | 90° view of input | output: completed point cloud

CVPR '17, Point Set Generation

Open problems

A better metric that takes the best of Chamfer and EMD?

How to add further structure constraints?

How to extend the pipeline to scene level?

How generalizable the method is?

In principle, what is the generalizability of a geometry estimator?
To what extend is 3D perception ability innate or learned?

Agenda

- Supervised Point Set Generation (cont)
- **Multidimensional Scaling**
- Parametric Shape Space for Homotopic Manifolds

Embedding / Sketching

- Definition: an embedding is a map $f{:}M \rightarrow H$ of a metric (M, d_M) into a host metric (H, ρ_H) such that for any $x,y \in M$:

$$d_M(x,y) \leq t\ \rho_H(f(x), f(y)) \leq D * d_M(x,y)$$

where D is the distortion (approximation) of the embedding f.

- Embeddings can be randomized: $\rho_H(f(x), f(y)) \approx d_M(x,y)$ with $1-\delta$ probability

- Types of embeddings:
  - From a norm (ℓ_1) into another norm (ℓ_∞)
  - From norm to the same norm but of *lower dimension* (dimension reduction)
  - From non-norms (Earth-Mover Distance, edit distance) into a norm
  - From given finite metric (shortest path on a planar graph) into a norm

[slide credit: Alexandr Andoni]

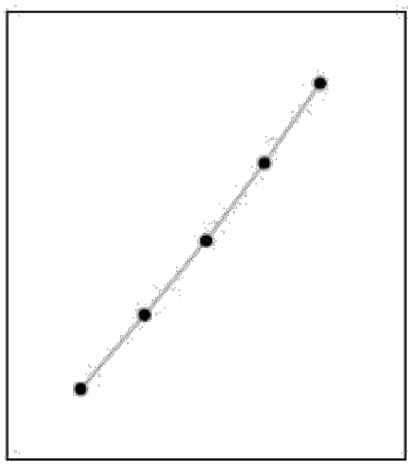
Distances and Dimensionality

- How do distances/dissimilarities determine dimensionality?

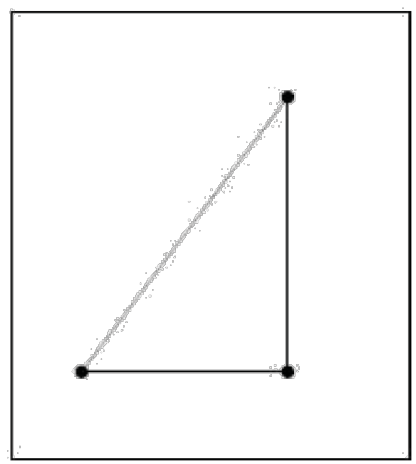
$$D = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 0 & 3 & 4 \\ 3 & 0 & 5 \\ 4 & 5 & 0 \end{bmatrix}$$

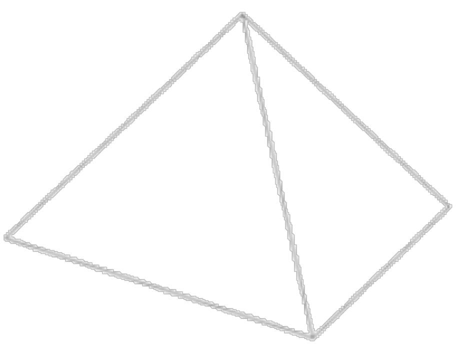
$$D = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$



k=1



k=2



k=3

Results for general metric space to l_∞

Theorem 2.7. *Every metric space embeds isometrically into ℓ_∞.*

Proof. We will prove this lemma only for finite metric spaces. Consider a metric space (X, d), where $X = (x_1, \ldots, x_n)$. It suffices to find a function $f : X \to \mathbb{R}_n$ such that (X, d) embeds isometrically into $(\mathbb{R}_n, \|\cdot\|)$. For $x_i \in X$ we define

$$f(x_i) = (d(x_1, x_i), d(x_2, x_i), \ldots, d(x_n, x_i))$$

Clearly it suffices to show for every $x_i, x_j \in X$ that $\|f(x_i) - f(x_j)\|_\infty = d(x_i, x_j)$. First we note that since d is a metric, it respects the $\triangle$-inequality, thus $d(x_i, x_k) - d(x_j, x_k) \leq d(x_i, x_j)$ for $k = 1, \ldots n$. It follows that

$$\max_k |d(x_i, x_k) - d(x_j, x_k)| \leq d(x_i, x_j),$$

or in other words

$$\|f(x_i) - f(x_j)\|_\infty \leq d(x_i, x_j). \quad (1)$$

On the other hand, the j-th coordinate of the vector $f(x_i) - f(x_j)$ is $d(x_j, x_i) - d(x_j, x_j) = d(x_i, x_j)$. Therefore the maximum coordinate of $f(x_i) - f(x_j)$ is at least $d(x_i, x_j)$ or in other words

$$\|f(x_i) - f(x_j)\|_\infty \geq d(x_i, x_j). \quad (2)$$

The lemma follows then from (1) and (2). □

http://www.cs.toronto.edu/~avner/teaching/S6-2414/LN1.pdf

Example results for planar EMD embedding

- Consider EMD on grid $[\Delta]\times[\Delta]$, and sets of size s

•

- Theorem [Cha02, IT03]: Can embed EMD over $[\Delta]^2$ into ℓ_1 with distortion $O(\log \Delta)$. Time to embed a set of s points: $O(s \log \Delta)$.

More: Sketching and Embedding are Equivalent for Norms

Results for Euc. space (dimension reduction)

Johnson–Lindenstrauss Flattening Lemma

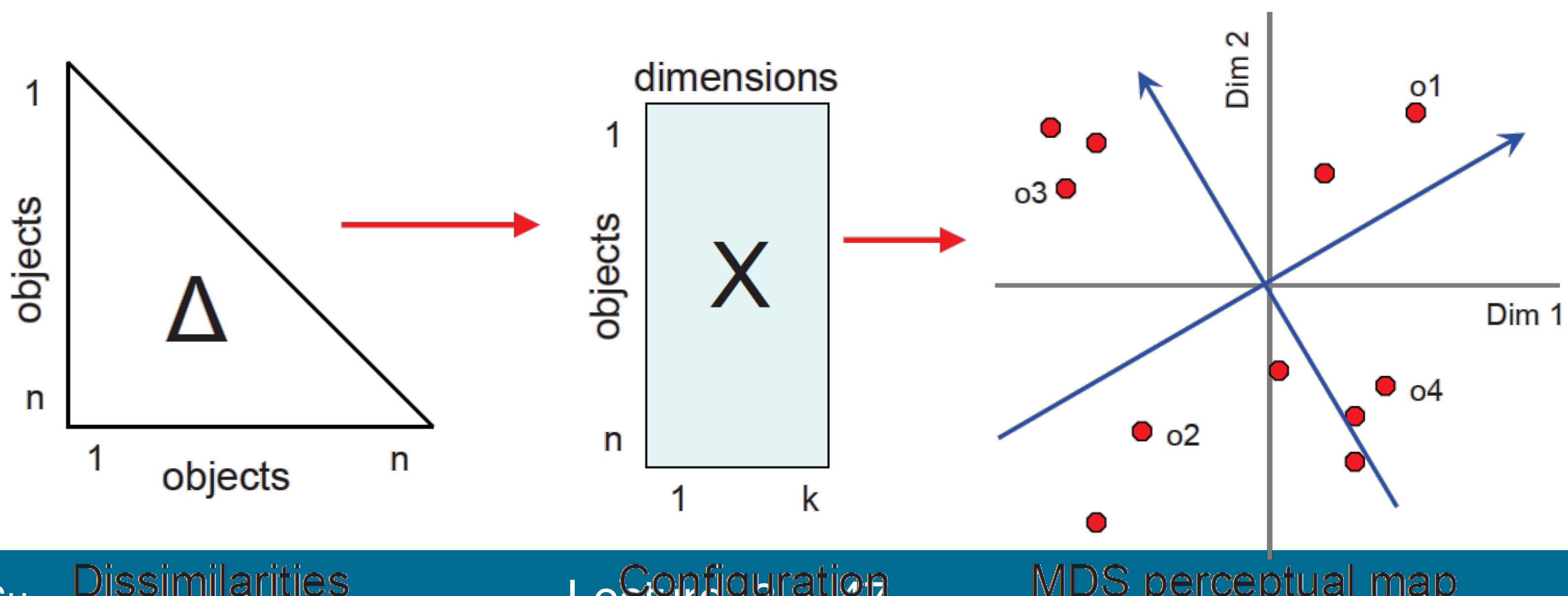
Given $0 < \varepsilon < 1$, a set X of m points in $\mathbb{R}^N$, and a number $n > 8\ln(m)/\varepsilon^2$, there is a linear map $f : \mathbb{R}^N \to \mathbb{R}^n$ such that

$$(1-\varepsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2$$

for all $u, v \in X$.

Multidimensional Scaling (MDS)

- A “distance preserving” embedding of the data into a Euclidean space
  - Sometimes distances are observed directly (e.g., similarity ratings)
  - Sometimes they can be calculated from a data table (e.g., Euclidean distances, correlations)



Formally …

- Given a (symmetric) matrix of pairwise "dis-similarities" between n objects / data sets

$$M = \left(\delta_{ij}\right)_{n \times n}$$

No need to satisfy the triangle inequality

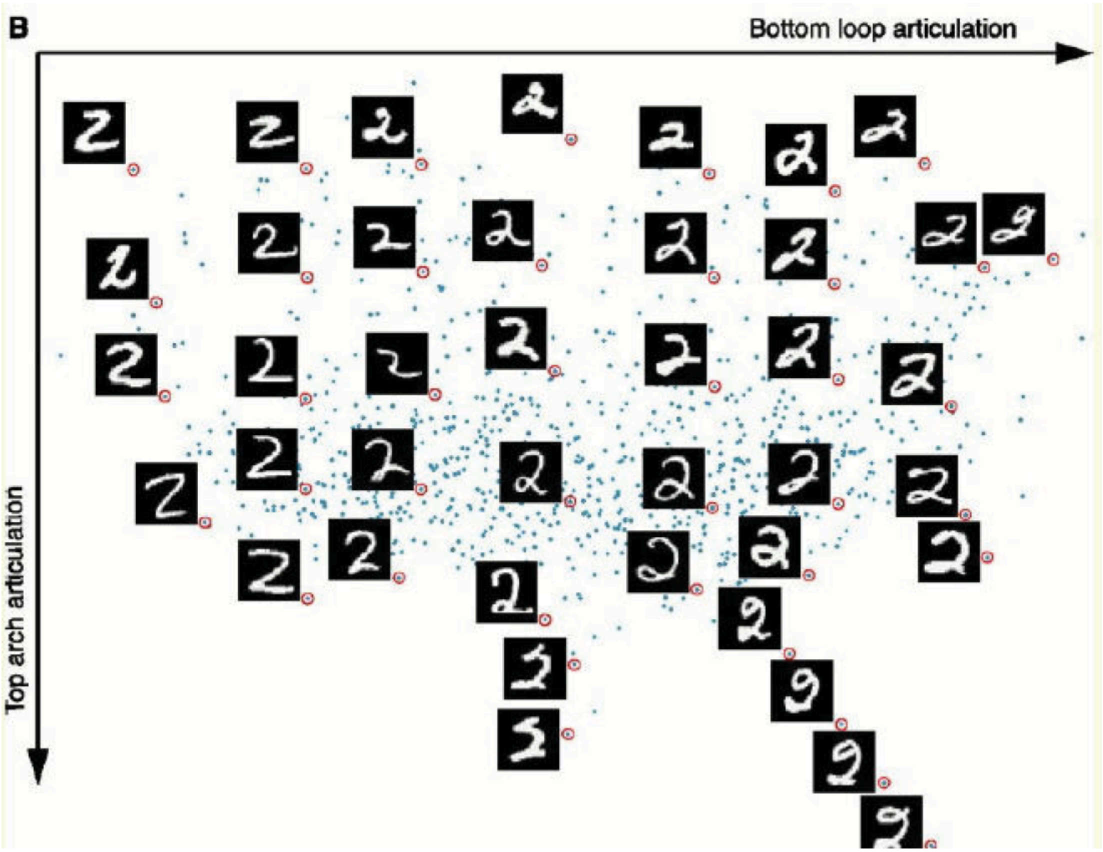
- Find n points in low-dimensional space R^d, so that their distance matrix is as close as possible to M

- Low d (=2,3) allows us to visualize the data directly

MDS Has Many Uses

- Psychology (perception, cognition)
- Political science (voting behavior, court decisions)
- Sociology (social network analysis)
- Archeology (artifact similarity)
- Biology/Chemistry (molecular structure, species analysis)
- Document retrieval & classification
- Graph layout
- Pattern recognition
- Dimension reduction
- …

Example: Pattern Recognition



MDS of judged similarity of handwritten "2"s

Goal: determine features important in pattern recognition

Classic Metric MDS

- Sometimes we can model our data as points in a high-dimensional Euclidean space – and we are looking for an embedding to a lower-dimensional space that preserves (absolute or relative) distances (in the high-d space) as much as possible.

- In this case the problem has a clean geometric solution.

Classic Metric MDS

- To go from dimension D down to dimension d
- Given data $X \in R^{D \times n}$

$$X = \begin{pmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{pmatrix} \quad \text{and} \quad M = \left(\text{dist}^2(\mathbf{x_i}, \mathbf{x_j})\right)_{n \times n}$$

- We look for X',

$$X' = \begin{pmatrix} | & & | \\ \mathbf{x_1}' & \cdots & \mathbf{x_n}' \\ | & & | \end{pmatrix} \in R^{d \times n}$$

- We can assume the $\mathbf{x}_i'$ are centered

Classic Metric MDS

- So that we minimize $\| M' - M \|$ (related to the *stress* of the system)

- where $M' = \left(\mathrm{dist}^2(\mathbf{x}_\mathrm{i}', \mathbf{x}_\mathrm{j}')\right) = \left(\left\|\mathbf{x}_\mathrm{i}' - \mathbf{x}_\mathrm{j}'\right\|^2\right) \in R^{n \times n}$

- M' is the Euclidean distances matrix for points $\mathbf{x}_\mathrm{i}'$.

The Math Details

- Ideally we want

$$M' = \left(\left\| \mathbf{x}_i' - \mathbf{x}_j' \right\|^2 \right) = M$$

$$\left(< \mathbf{x}_i' - \mathbf{x}_j', \; \mathbf{x}_i' - \mathbf{x}_j' > \right) = M$$

$$\left(\| \mathbf{x}_i' \|^2 + \| \mathbf{x}_j' \|^2 - 2 < \mathbf{x}_i', \mathbf{x}_j' > \right) = M$$

$$\begin{pmatrix} \|\mathbf{x}_1'\| & \|\mathbf{x}_1'\| & \cdots & \|\mathbf{x}_1'\| \\ \|\mathbf{x}_2'\| & \|\mathbf{x}_2'\| & \cdots & \|\mathbf{x}_2'\| \\ & & \vdots & \\ \|\mathbf{x}_n'\| & \|\mathbf{x}_n'\| & \cdots & \|\mathbf{x}_n'\| \end{pmatrix} \quad \begin{pmatrix} \|\mathbf{x}_1'\| & \|\mathbf{x}_2'\| & & \|\mathbf{x}_n'\| \\ \|\mathbf{x}_1'\| & \|\mathbf{x}_2'\| & & \|\mathbf{x}_n'\| \\ \vdots & \vdots & \cdots & \vdots \\ \|\mathbf{x}_1'\| & \|\mathbf{x}_2'\| & & \|\mathbf{x}_n'\| \end{pmatrix}$$

want to get rid of these

$$\underbrace{\begin{pmatrix} \text{---} & \mathbf{x}_1' & \text{---} \\ & \vdots & \\ \text{---} & \mathbf{x}_n' & \text{---} \end{pmatrix}}_{X'^T} \underbrace{\begin{pmatrix} | & & | \\ \mathbf{x}_1' & \cdots & \mathbf{x}_n' \\ | & & | \end{pmatrix}}_{X'}$$

The Magic Matrix J

$$J = \begin{pmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ \vdots & & \ddots & \vdots \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} \end{pmatrix}_{n\times n} = I - \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 1 & 1 \end{pmatrix} = I - \frac{1}{n}K$$

$$\begin{pmatrix} a & a & \cdots & a \end{pmatrix} \cdot J = 0$$

$$J \cdot \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = 0$$

So We Get to The Gram Matrix

Cleaning the system:

$$\times J \Big/ \begin{pmatrix} \|\mathbf{x}_1'\| & \|\mathbf{x}_1'\| & \dots & \|\mathbf{x}_1'\| \\ \|\mathbf{x}_2'\| & \|\mathbf{x}_2'\| & \dots & \|\mathbf{x}_2'\| \\ & & \vdots & \\ \|\mathbf{x}_n'\| & \|\mathbf{x}_n'\| & \dots & \|\mathbf{x}_n'\| \end{pmatrix} + \begin{pmatrix} \|\mathbf{x}_1'\| & \|\mathbf{x}_2'\| & & \|\mathbf{x}_n'\| \\ \|\mathbf{x}_1'\| & \|\mathbf{x}_2'\| & & \|\mathbf{x}_n'\| \\ \vdots & \vdots & \dots & \vdots \\ \|\mathbf{x}_1'\| & \|\mathbf{x}_2'\| & & \|\mathbf{x}_n'\| \end{pmatrix} - 2X'^T X' = M \Big/ \times J$$

$$-2X'^T X' = JMJ$$

$$X'^T X' = -\tfrac{1}{2} JMJ =: B$$

Note that $X'K = KX'^T = 0$, as X' is centered

$$\boxed{X'^T X' = B}$$

So from the distance matrix we can get the Gram (inner product) matrix.

And Finally the Spectral Hammer

We will use the spectral decomposition of B:

$$X'^T X' = B = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix}^T$$

$$X'^T X' = \underbrace{\underbrace{\begin{pmatrix} | & & | & \vdots & | \\ | & & | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_d & \vdots & \mathbf{v}_n \\ | & & | & & | \\ | & & | & \vdots & | \end{pmatrix}}_{n \times d} \begin{pmatrix} \underbrace{\begin{matrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_d} \end{matrix}}_{d \times d} & \\ & \ddots \\ & & \sqrt{\lambda_n} \end{pmatrix}}_{X'^T} \underbrace{\begin{pmatrix} \sqrt{\lambda_1} & & & & \\ & \ddots & & & \\ & & \sqrt{\lambda_d} & & \\ & & & \ddots & \\ & & & & \sqrt{\lambda_n} \end{pmatrix}^T \begin{pmatrix} | & & | & \vdots & | \\ | & & | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_d & \vdots & \mathbf{v}_n \\ | & & | & & | \\ | & & | & \vdots & | \end{pmatrix}^T}_{X'}$$

So We Get the X'

So we find X' by throwing away the last $n-d$ eigenvalues

$$X' = \begin{pmatrix} \text{—} & \sqrt{\lambda_1}\,\mathbf{v_1} & \text{—} \\ \cdots & \cdots & \cdots \\ \text{—} & \sqrt{\lambda_d}\,\mathbf{v_d} & \text{—} \end{pmatrix}_{d \times n}$$

For this X' :

$$X' = \arg\min_{X'} \left\| X'^T X' - B \right\|_{L^2}$$

$$\|A\|_{L^2} = \sqrt{\sum_{i,j} A_{ij}^2}$$

This choice minimizes the
inner product (and distance) loss

More General Metric MDS

- In general, we minimize directly the square loss on distances

$$\text{stress} = \mathcal{L}(\hat{d}_{ij}) = \left(\sum_{i<j} (\hat{d}_{ij} - f(d_{ij}))^2 / \sum d_{ij}^2 \right)^{\frac{1}{2}}$$

- Sammon mapping

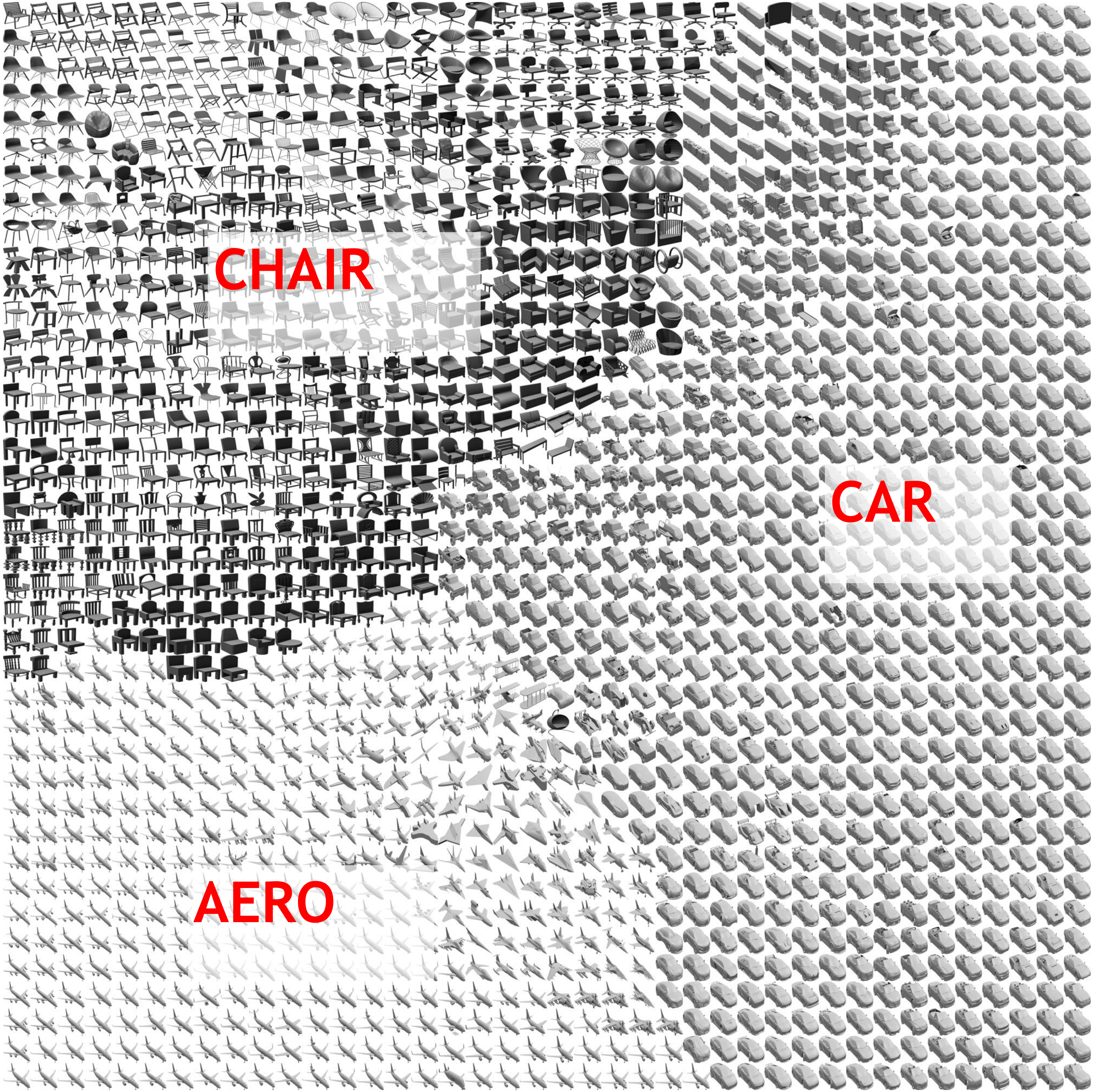
$$\text{Sammon's stress}(\hat{d}_{ij}) = \frac{1}{\sum_{\ell<k} d_{\ell k}} \sum_{i<j} \frac{(\hat{d}_{ij} - d_{ij})^2}{d_{ij}}$$

  - This weighting system normalizes the squared-errors in pairwise distances by using the distance in the original space. As a result, Sammon mapping preserves the small d_{ij}, giving them a greater degree of importance in the fitting procedure than for larger values of d_{ij}

Generally solved by gradient descent

Agenda

- Supervised Point Set Generation (cont)
- Multidimensional Scaling
- **Parametric Shape Space for Homotopic Shapes**


CHAIR
CAR
AERO

Every point in the shape space is a "valid shape"?

[Wu et al, Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling]

Homotopy

For continuous functions f and g from a topological space X to a topological space Y:

- f and g are homotopic iff here exists a continuous function $H : X \times [0,1] \rightarrow Y$, *such that*

$$H(x,0) = f(x) \text{ and } H(x,1) = g(x).$$

Homotopy

Intuition: To construct the family of deformable shapes (face, body, etc.)