IMPERIAL COLLEGE LONDON

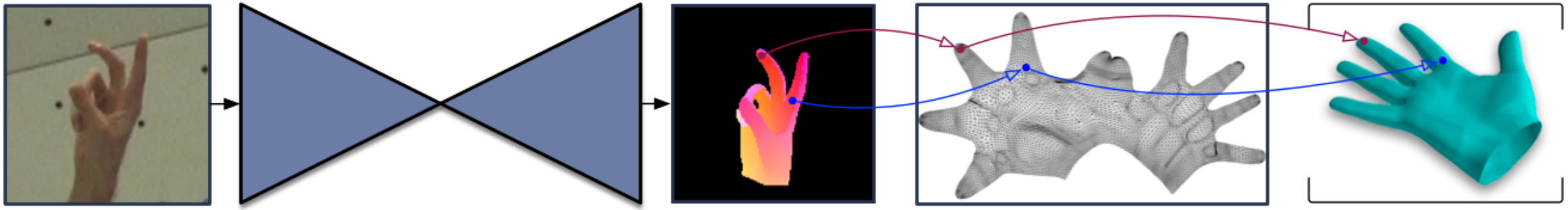
Dense Hand Pose Estimation

MRes Individual Project Presentation
12 September 2018

Dominik Kulon
Author

Dr. Stefanos Zafeiriou
Supervisor

Dense Hand Pose Estimation



Dense Regression Network

Applications

- Human motion tracking
- Animating digital characters
- Human-computer interaction
- Recognizing body language
- Recognizing human activities
- Generating body models

Challenges

Pose Space

- A hand has 27 degrees of freedom.
- Deep learning requires exhaustive datasets.

Occlusions

- Fingers are similar to each other.
- Self-occlusions of joints are common.
- Occlusions can be caused by grasped objects.

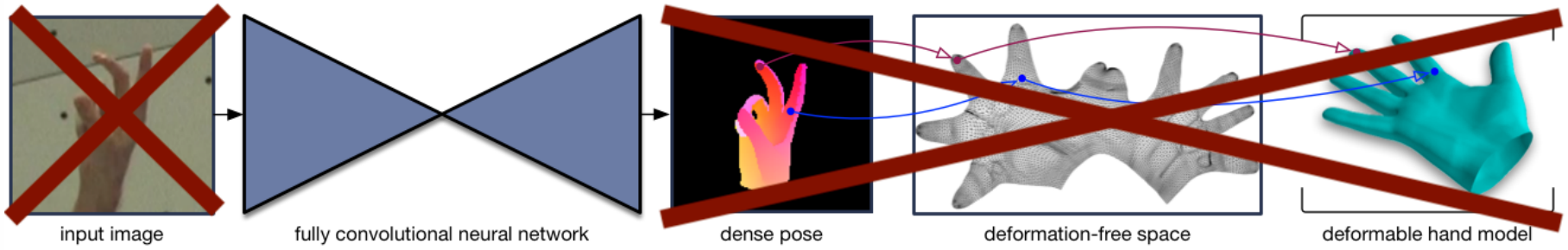
Image Annotations

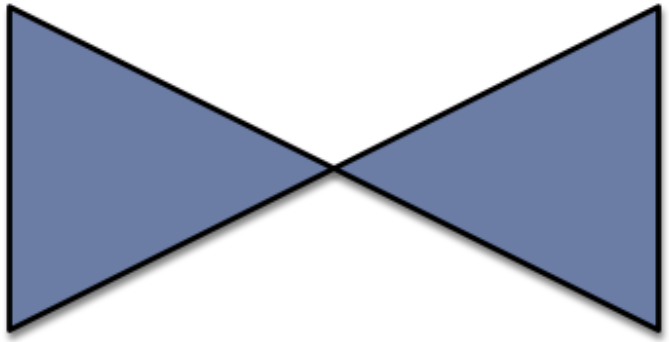
- The hand covers a small area in the image.
- Manual annotations are difficult and ambiguous.

Proposed System

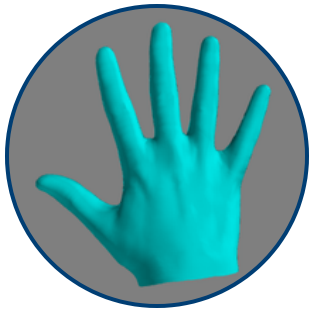
- No adequate hand model.
- No densely annotated training data.

Starting Point





Implementation Steps



Hand Model

Create a statistical deformable hand model with shape deformations learned from hand scans.

Model Fitting

Fit the model into spare annotations of RGB images to generate the ground truth data.

Dense Shape Regression

Train a dense hand pose estimation system.

Task 1

Hand Model

Definition

$$M(\boldsymbol{\beta}, \boldsymbol{\theta}; \phi) : \mathbb{R}^{|\boldsymbol{\beta}| \times |\boldsymbol{\theta}|} \rightarrow \mathbb{R}^{3N}$$

- **N - number of vertices**
- **beta - shape parameters**

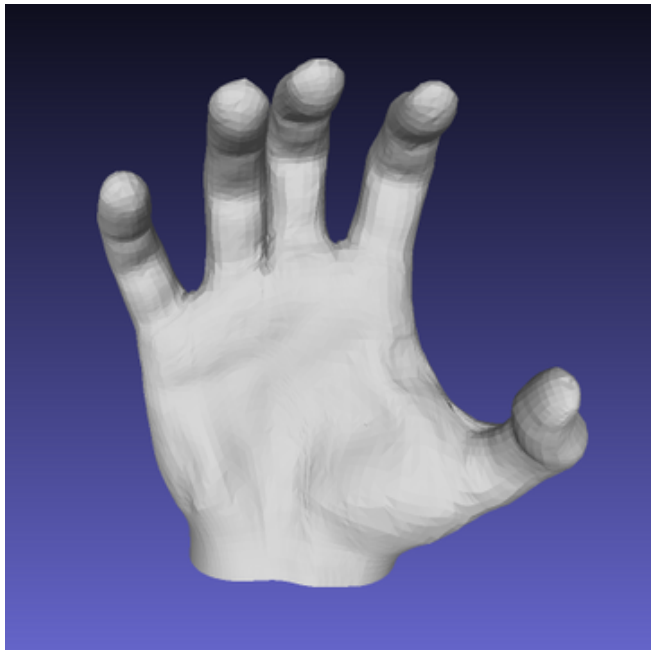
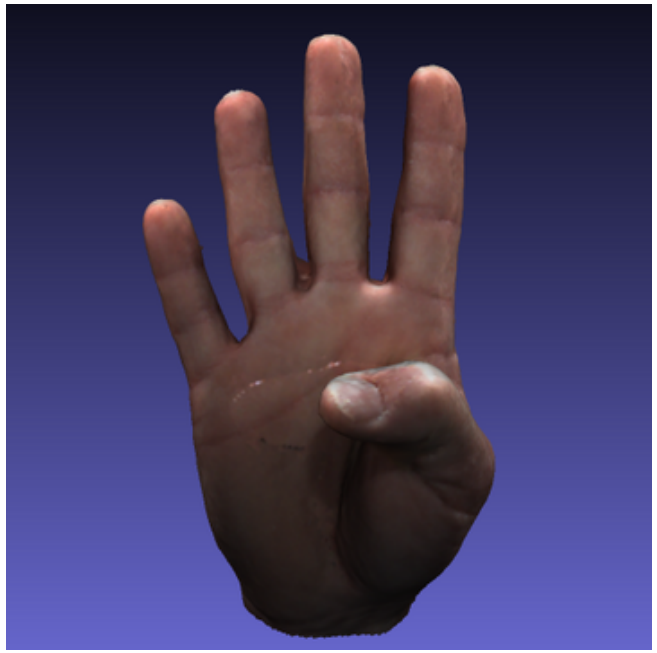
  Shape: Low-Dimensional Embedding
- **theta - pose parameters**

  Pose: Linear Blend Skinning
- **phi - a set of learned parameters**

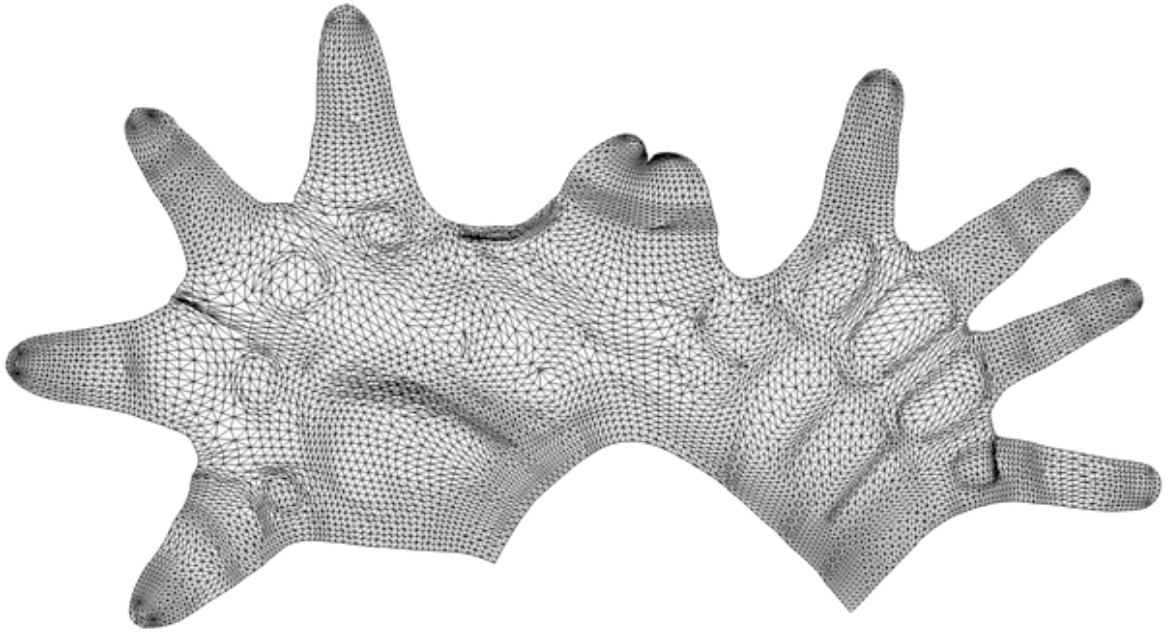
  principal components, skinning weights, pose-dependent displacements...

Registrations



UV Map



- A bijective map from the 3D hand model to a two-dimensional space.
- (u, v) coordinates

Demo

Task 2

Model Fitting

Pose and Shape Optimization

$$\{\beta^*, \theta^*\} = arg\min_{\beta,\theta}(E_Y + E_{prior} + E_{reg})$$

- **Reconstruction Error Term**

  Penalizes the difference between the model joints/vertices and data annotations.

- **Prior Error Term**

  Addresses pose ambiguity issues with 2D annotations.

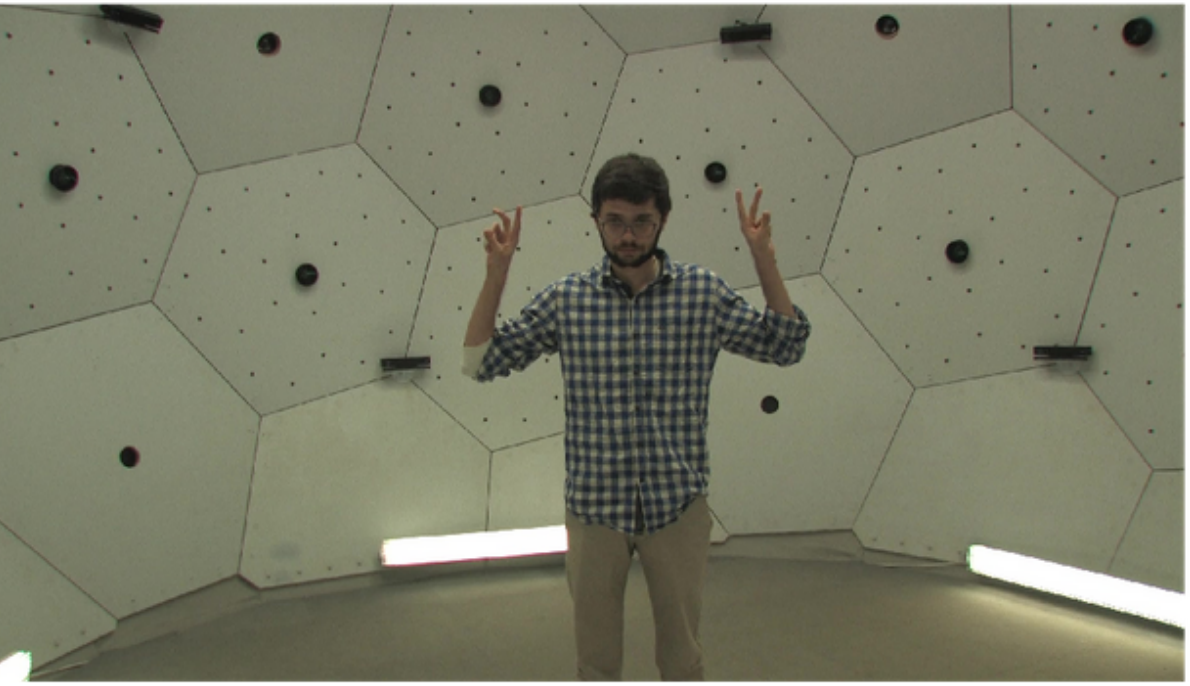
- **Regularization Error Term**

  Regularizes the optimization procedure to ensure realistic deformations and matrix sparsity.

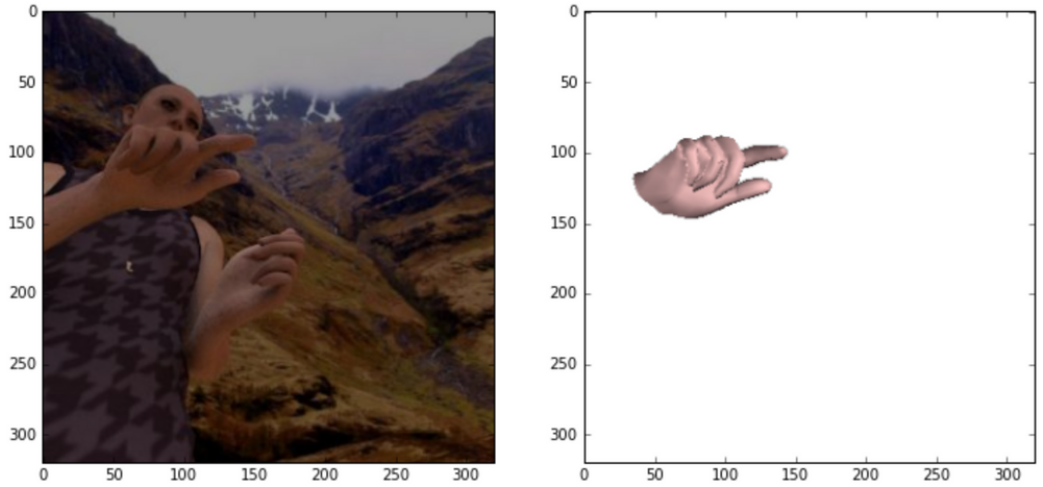
Dataset Generation

- Panoptic DomeDb - 30 synchronized HD cameras.
- Includes camera parameters and 3D annotations.
- We take 25,000 video frames.
- We use both hands.
- 30 x 25,000 x 2 = 1,500,000 samples

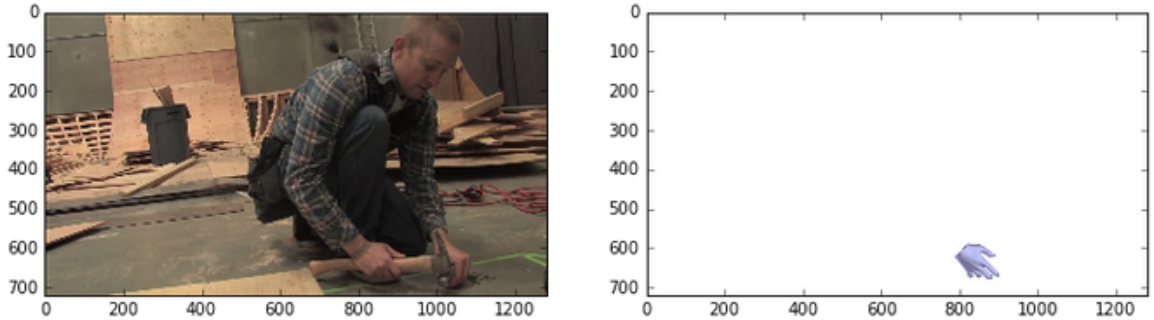
Model Fitting



Pose Ambiguity with 2D Annotations



Rendered Hand Pose Dataset

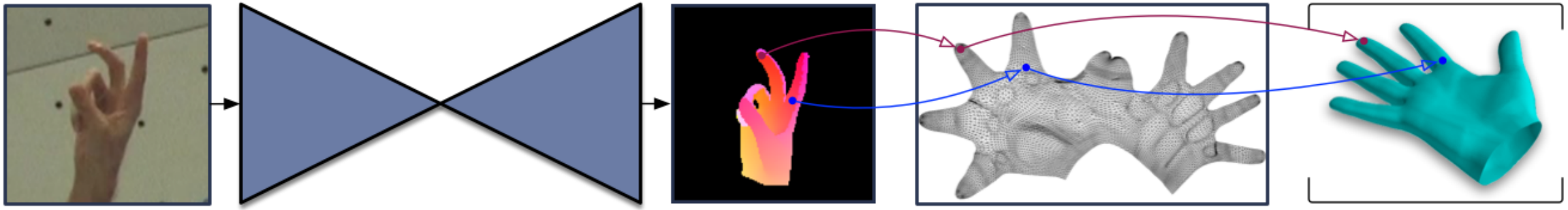


HandDB

Task 3

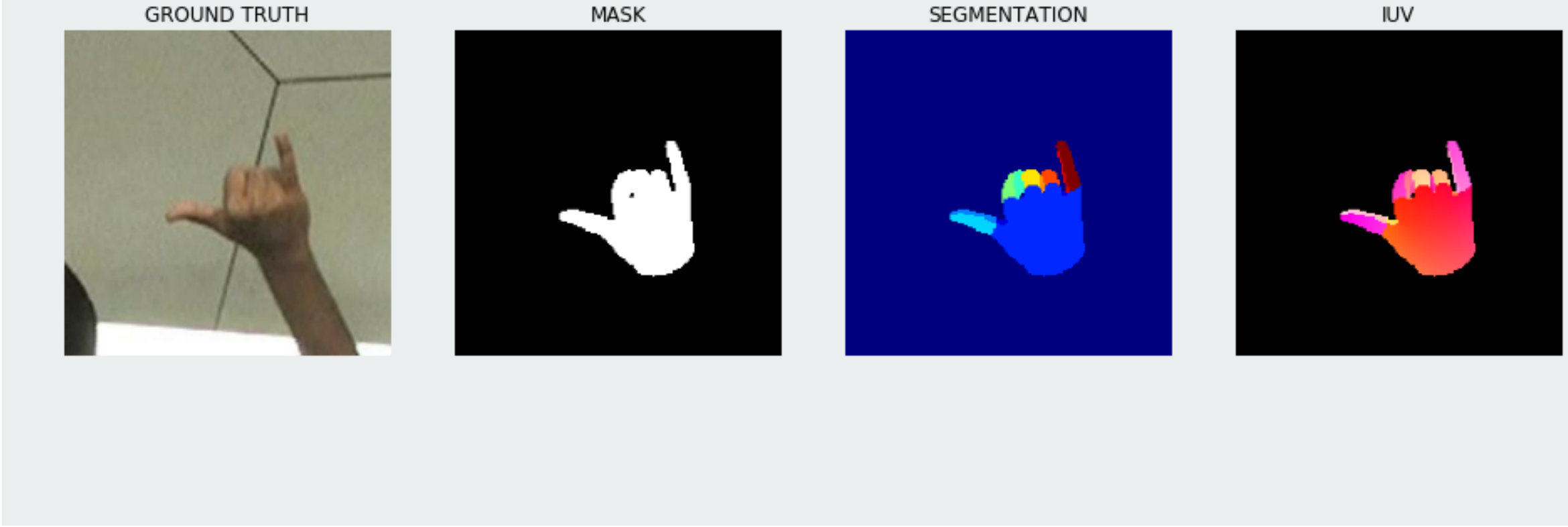
Dense Shape Regression

Recap



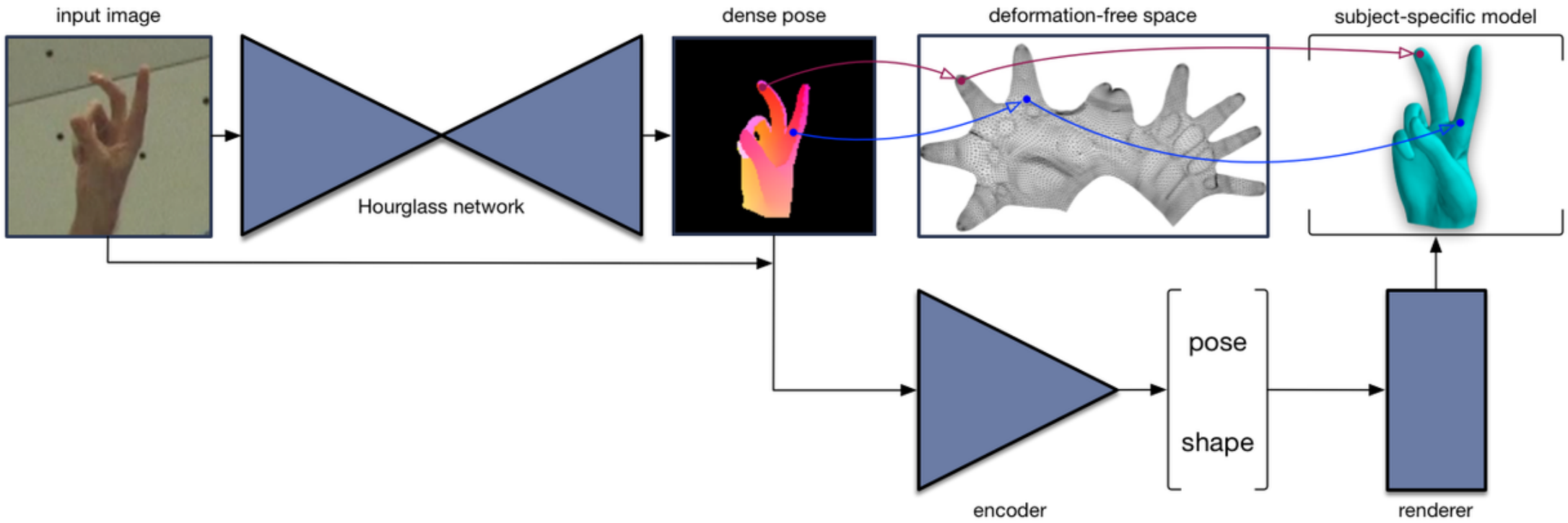
Dense Regression Network

Training Data

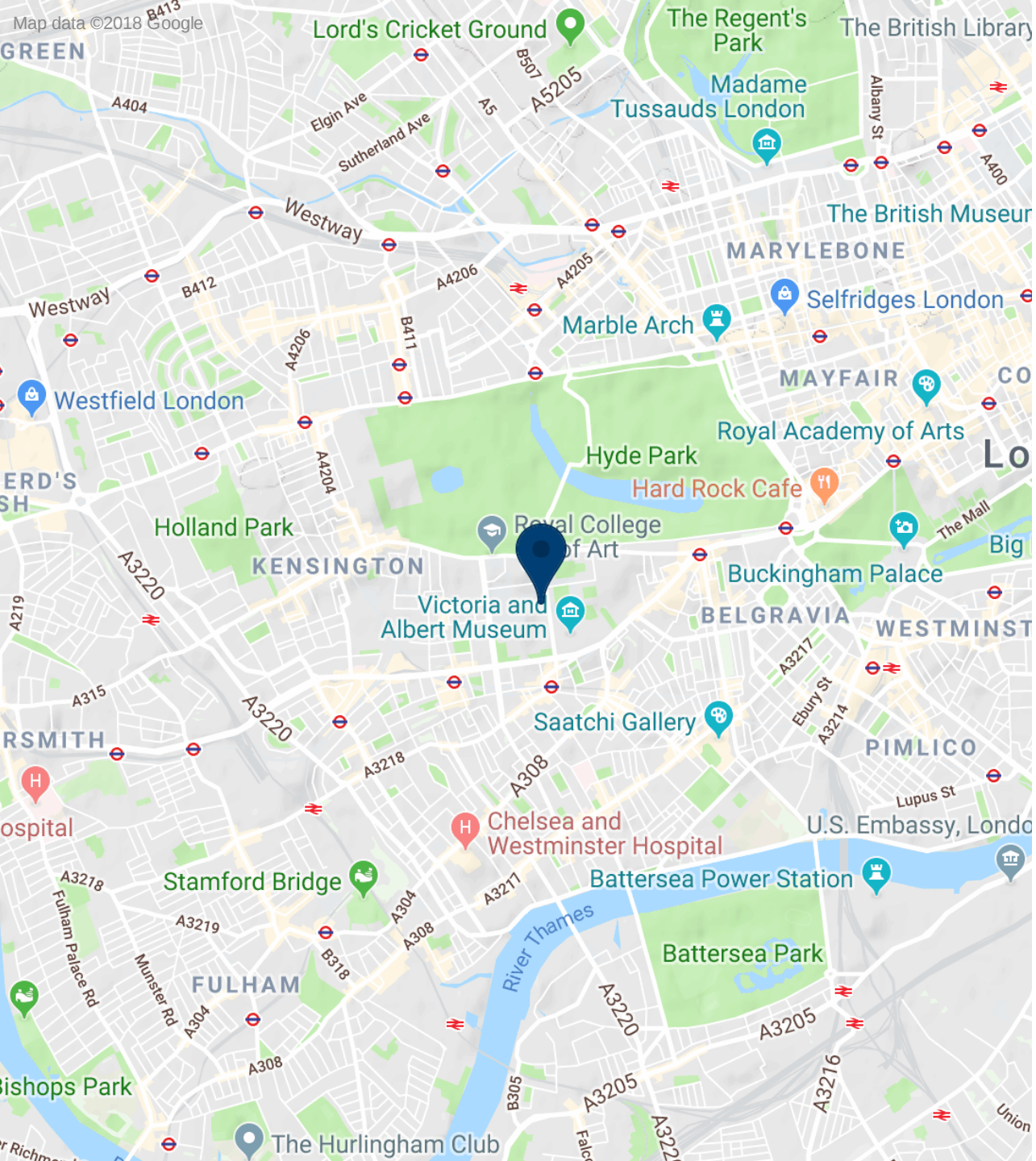


Future Work

Neural Network Idea



Dense Hand Pose Renderer



Department of Computing

Imperial College London

imperial.ac.uk/computing