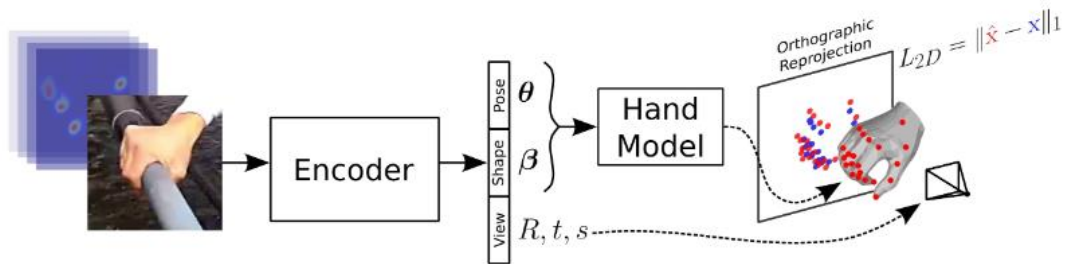**3d hand shape and pose from Images in the wild (2019)**

Goal:

- Operate on: in the wild images
- Predict 3d hand shape and pose

Dataset

- Synthetic dataset to pretrain encoder:

  - o Use decoder, sample poses, shapes, apply rotations, scaling, translation
  - o Use 3d scans to generate hand model, render on top of random background images
- Training:
  - o Panoptic (15000 images) + MPII (2000) + STEREO (15000)
- Evaluation:
  - o Dexter / Object Dataset
  - o EgoDexter

Main obstacles:

- Lack of large datasets with reliable ground truth
- Current datasets (3d annotated) not able to generalize network well enough for in the ild pics

MANO:

- Use this model
- Handmodel parameterized by angle of rotations specified for each joint
- 16 joints + 5 for the fingertips

Network

Input:

- Optional: 2D joints detection heatmaps (from independent CNN)
- Monocular RGB images

Encoder:

- Predicts a set of hand view parameters
    o Hand shape
    o Pose parameters
    o Set of view parameters
- Deep convolutional encoder
- Pre-trained with synthetic examples (self-created)
- Use res-net50 network, adjust final fully connected layer

Decoder:

- No training required
- Generate a triangulated 2d mesh and its underlying skeleton
- 2 parts:
    o Pre-computed articulated mesh deformation hand model that generates a 3d triangulated mesh from the hand parameters (+ underlying skeleton)
    o Re-projection module controlled by the view parameters (project generated hand into image domain)
        ▪ Use rotation, translation and scale matrix
        ▪ Project orthogonally

Training / loss:

- End to end using 2d and 3d joint annotations without supervision on the hand / view params, except for a regularization on the hand params to ensure their magnitude is small
- Combine multiple losses:
    o 2d joint reprojection loss (
    o 3d joint loss
    o Hand mask loss (binary mask, occlusion aware, 2d)
        ▪ Ground truth generated by GrabCut algo
    o Model parameter regularization loss (hand model params)

Post processing:

- none