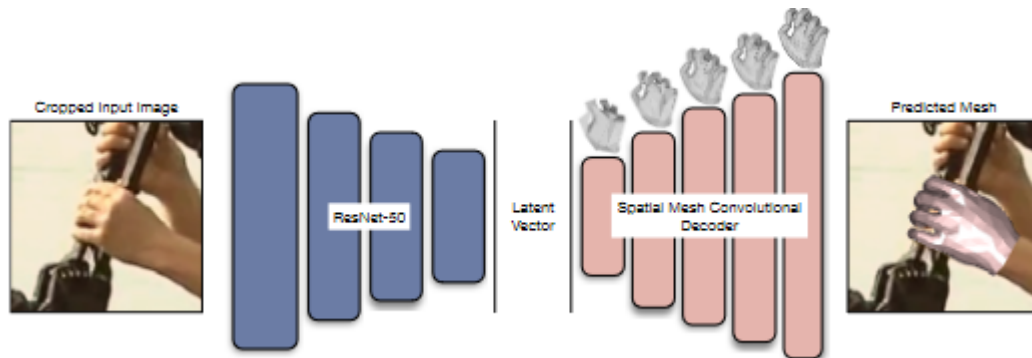**Weakly Supervised Mesh Convolutional Hand Reconstruction in the Wild (2020)**

https://www.arielai.com/mesh_hands/

Goal:

- Operate on: in the wild images

MANO

- Handmodel parameterized by angle of rotations specified for each joint

Dataset

- Generate their own dataset using youtube videos
    - Sign language
- Use weakly supervised approach:
    - First detect keypoints using open-pose
    - Lift them into 3d space by iteratively fitting a 3d deformable hand model
    - (thourgh sparse matrix that regresses from model vertices to 16 joints and 5 fingertip positions)a
    - Fit model to 2d annotations
- Fit MANO model in tandem with prior on angles of rotation
    - + camera scaling / translation
- 778 vertices, 16 joints
- Constrain joint angles to lie in convex hull of precomputed cluster centers
- More Datasets available:
    - MPII
    - RHD
    - FreiHAND
    - MTC
    - STB

Network

Input:

- Monocolar, RGB Picture, Cropped 192x192px, hand-centered

Encoder:

- ResNet-50,

Latent Vector

- 64 parameters

Decoder:

- Decode mesh directly from image encoding and apply spatial convolutions with pooling layers
- ReLu activation function
- Spiral filters to generate hands directly from an image encoding
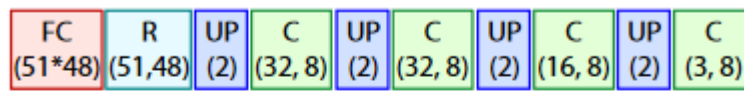 - Spiral patch operator for constructing spatial neighbourhoods

FC	R	UP	C	UP	C	UP	C	UP	C
(51*48)	(51,48)	(2)	(32, 8)	(2)	(32, 8)	(2)	(16, 8)	(2)	(3, 8)

Figure 3: Architecture of the spiral decoder. $C(w, l)$: A spiral convolution with w filters and a spiral length l; $U(p)$: upsampling by a factor p; $FC(d)$: a fully connected layer with d output dimension; $R(d_1, d_2)$: vector reshaping into a $d_1 \times d_2$ matrix.

Loss:

- On meshes with points localized in the image coordinate system
- Leveraging spatial mesh convolutions