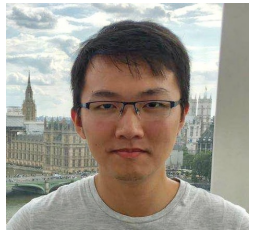
Few-shot Geometry-Aware Keypoint Localization

Paper tag: THU-PM-070


Xingzhe He


Gaurav Bharaj


David Ferman


Helge Rhodin


Pablo Garrido

THE UNIVERSITY OF BRITISH COLUMBIA

FLAWLESS

Brief Introduction

Goal: Train a keypoint detector with

Many unannotated examples

+

Few annotated examples

Method:

1. Adapt existing 2D unsupervised keypoints methods
2. Constrain the keypoints in 3D, and model keypoint occlusion (uncertainty)

Results:

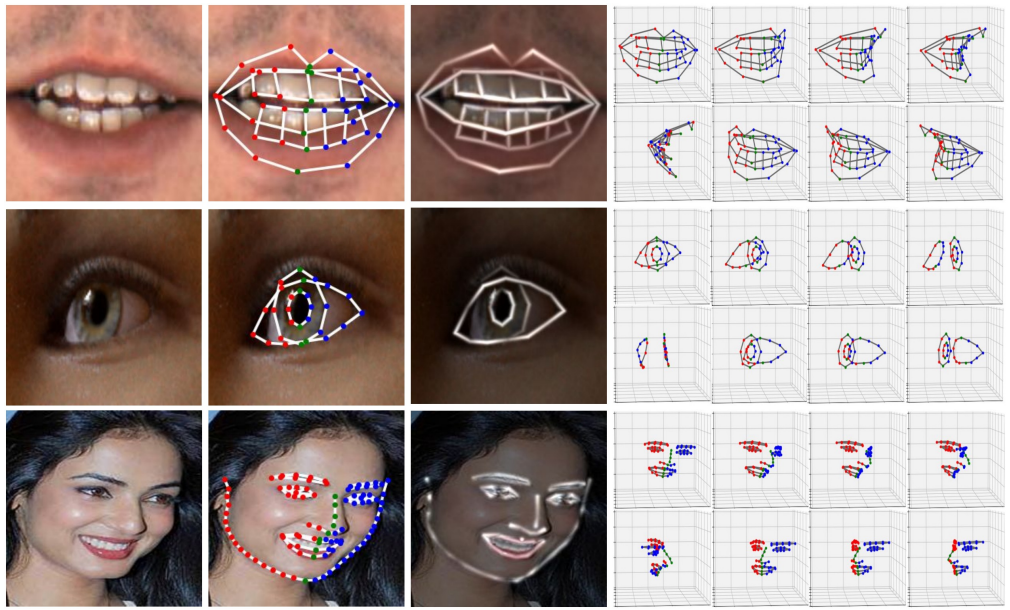


Image Keypoints Uncertainty Structure

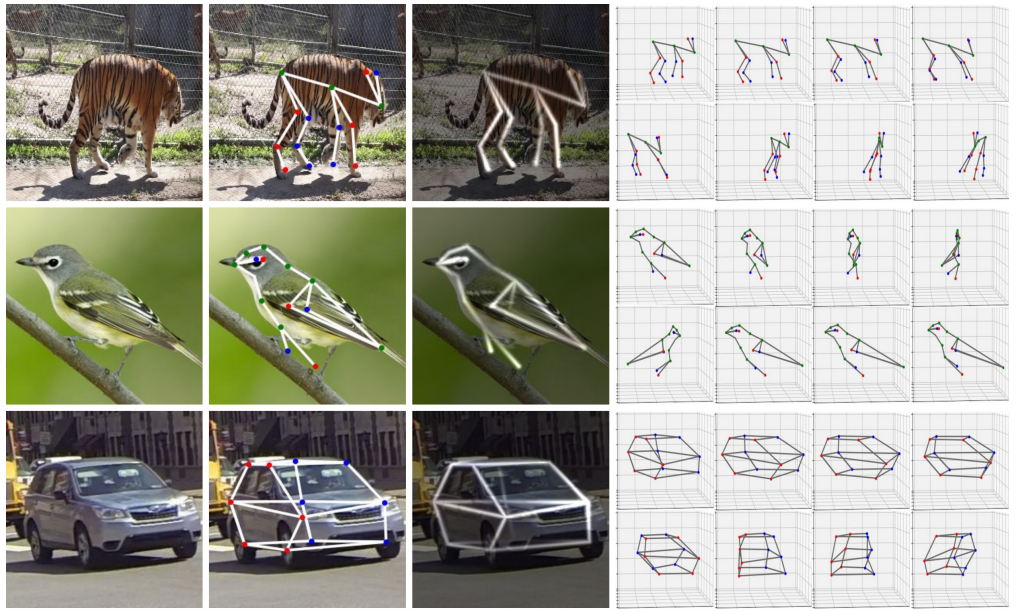
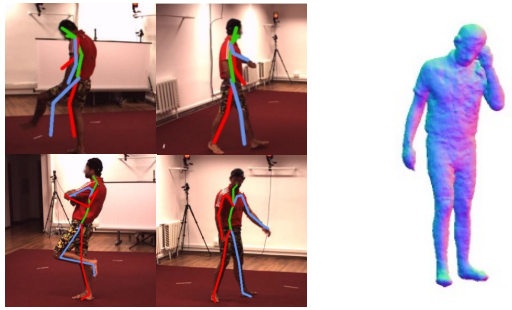


Image Keypoints Uncertainty Structure

Motivation

Keypoint is a common middle representation that are widely used in high-level tasks, such as

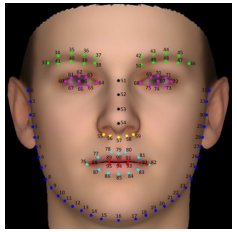
Pose transfer



3D reconstruction

However,

- Supervised keypoint annotation is expensive and tedious.
- Unsupervised keypoints are not human interpretable
- Semi-supervised methods still need more than thousands of annotated examples
- Existing few-shot methods only work on specific area, e.g., faces, X-rays.



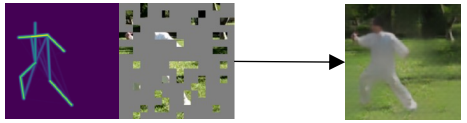
Human annotated keypoints

Unsupervised keypoints

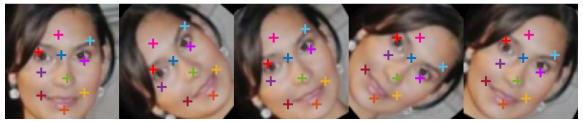
Idea

Basic idea: Inject the few-shot supervision to unsupervised keypoint detection

Image reconstruction from edge maps and masked images (He' 2022 NeurIPS)



Keypoint follow the same transformation applied to the image (Thewlis' 2017 ICCV)



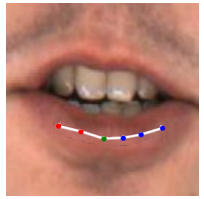
In each batch, we sample partially from annotated examples, and supervise their learned keypoints

2D keypoint supervision ←

→ Unsupervised Learning
Image reconstruction
Equivariant transformation

Idea

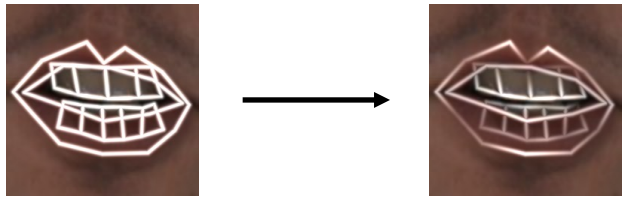
However, human annotated keypoints are projection from 3D, which contains occlusion.



Here the teeth keypoints are occluded, behind the lips

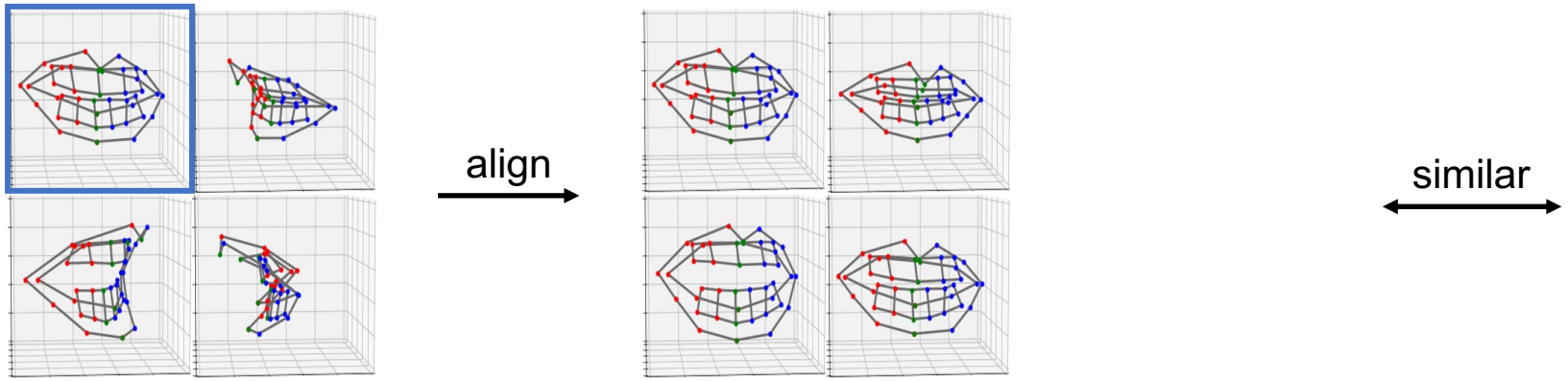
Constrain keypoints in 3D:

- Model occlusion (uncertainty)



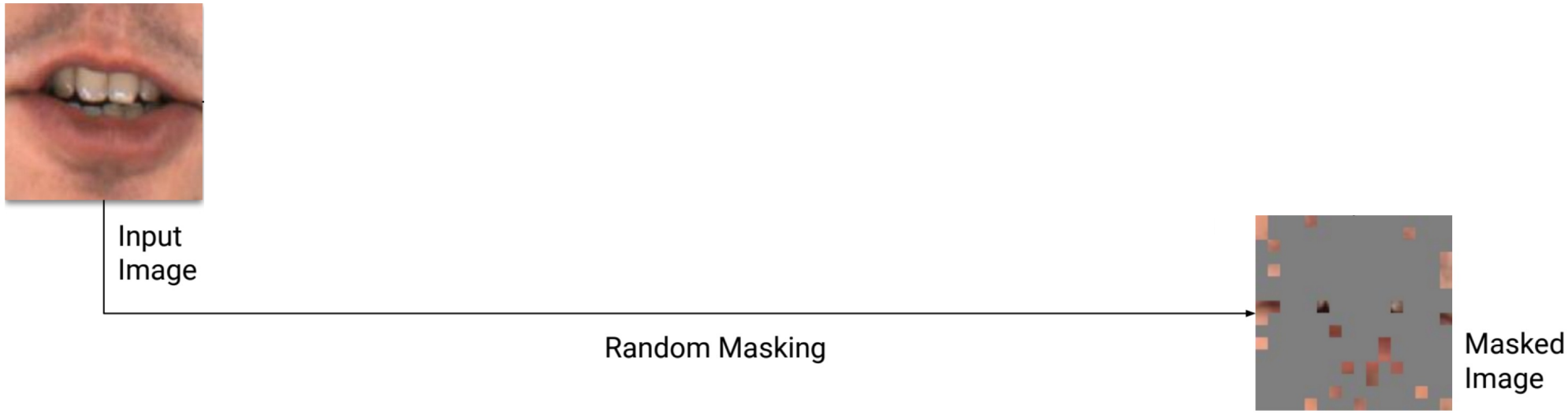
- Similarity in 3D

For each example in a batch

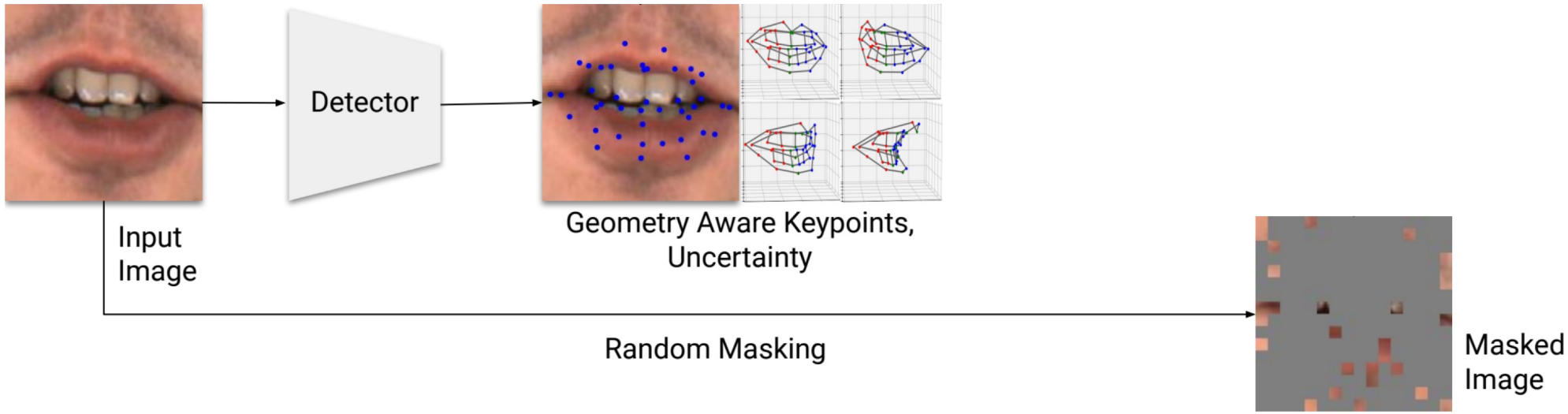


Constrain in parts: constrain upper teeth, bottom teeth, lips, and whole object separately

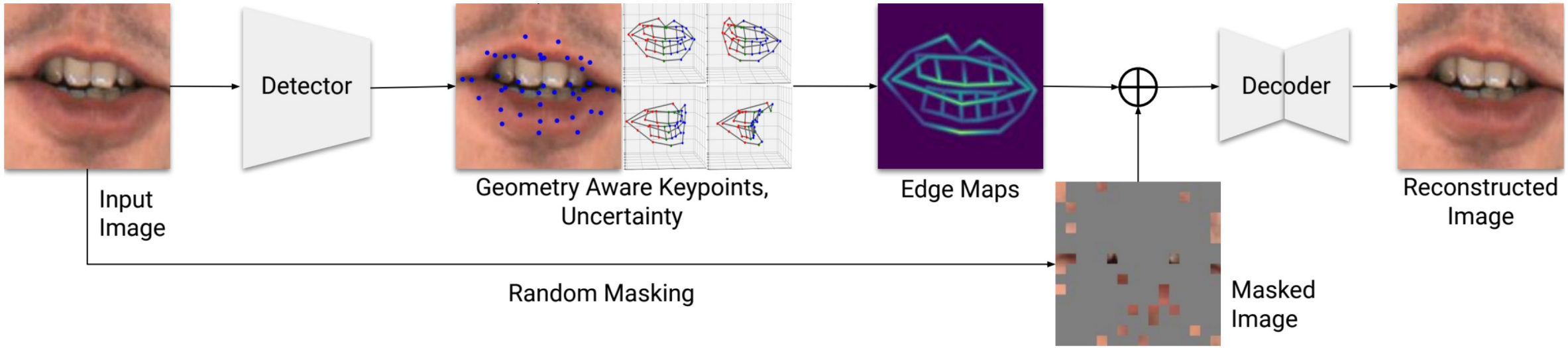
Overview



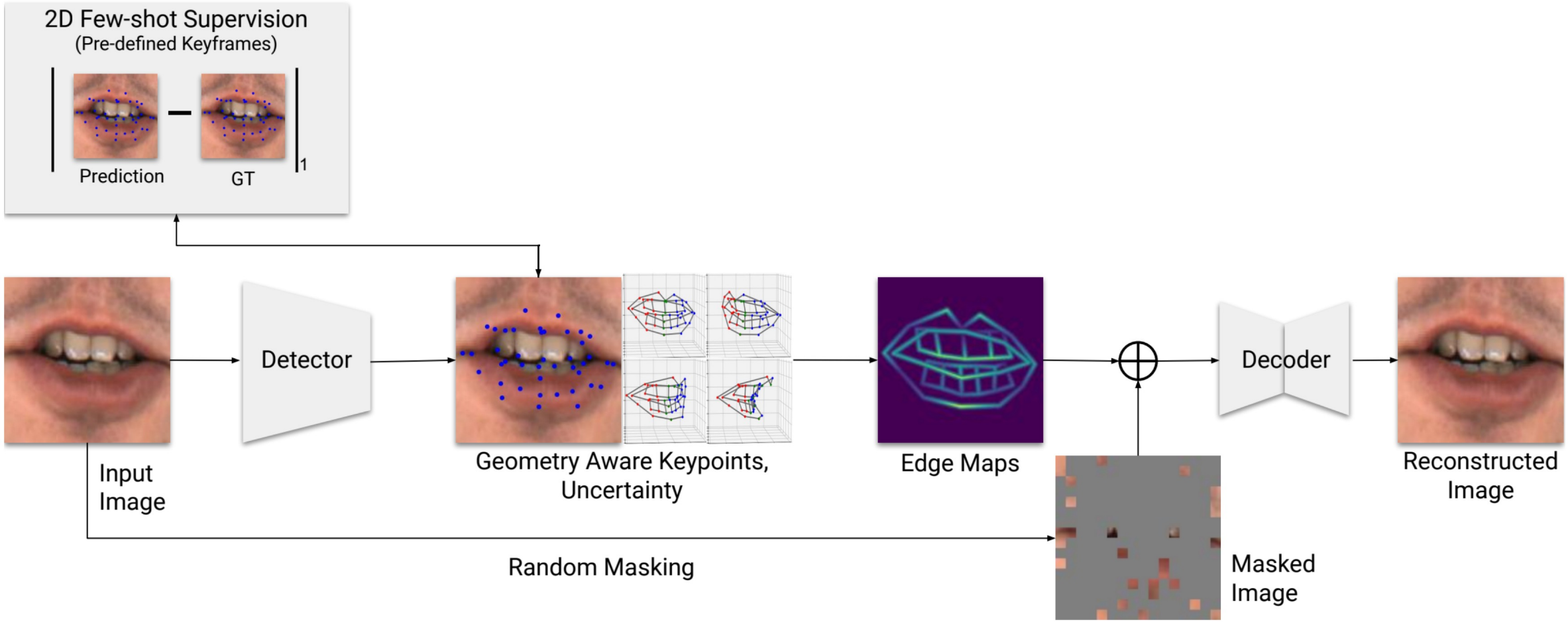
Overview


Detector
Input
Image
Geometry Aware Keypoints,
Uncertainty
Random Masking
Masked
Image

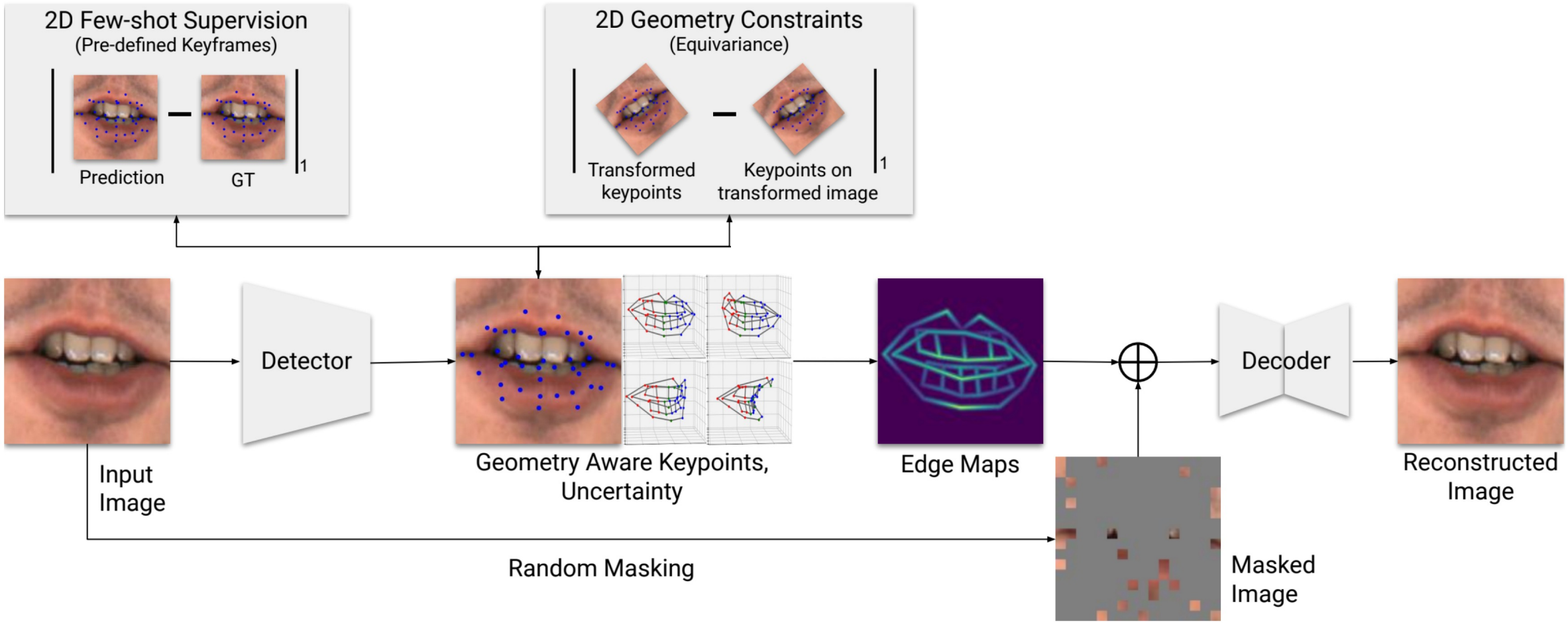
Overview


Detector
Input
Image
Geometry Aware Keypoints,
Uncertainty
Edge Maps
Decoder
Reconstructed
Image
Random Masking
Masked
Image

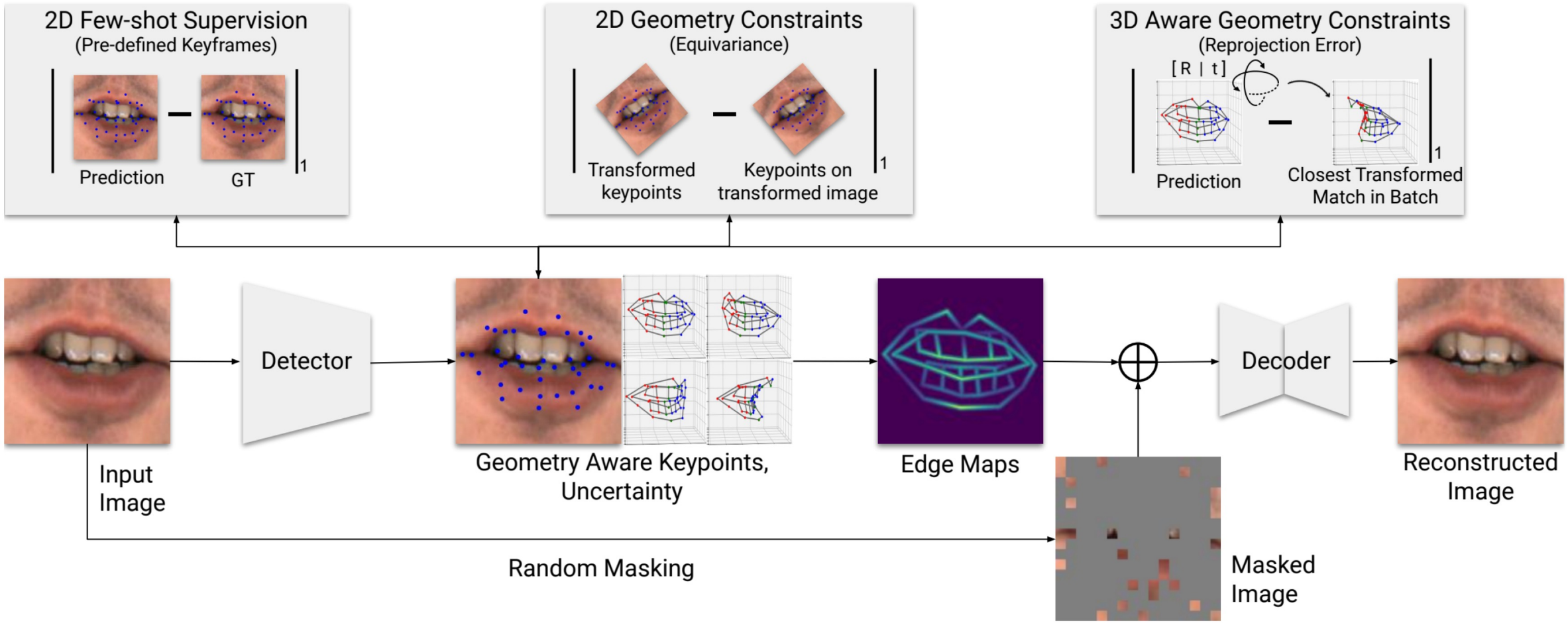
Overview


2D Few-shot Supervision
(Pre-defined Keyframes)
Prediction
GT
1
Detector
Geometry Aware Keypoints,
Uncertainty
Edge Maps
Decoder
Reconstructed
Image
Input
Image
Random Masking
Masked
Image

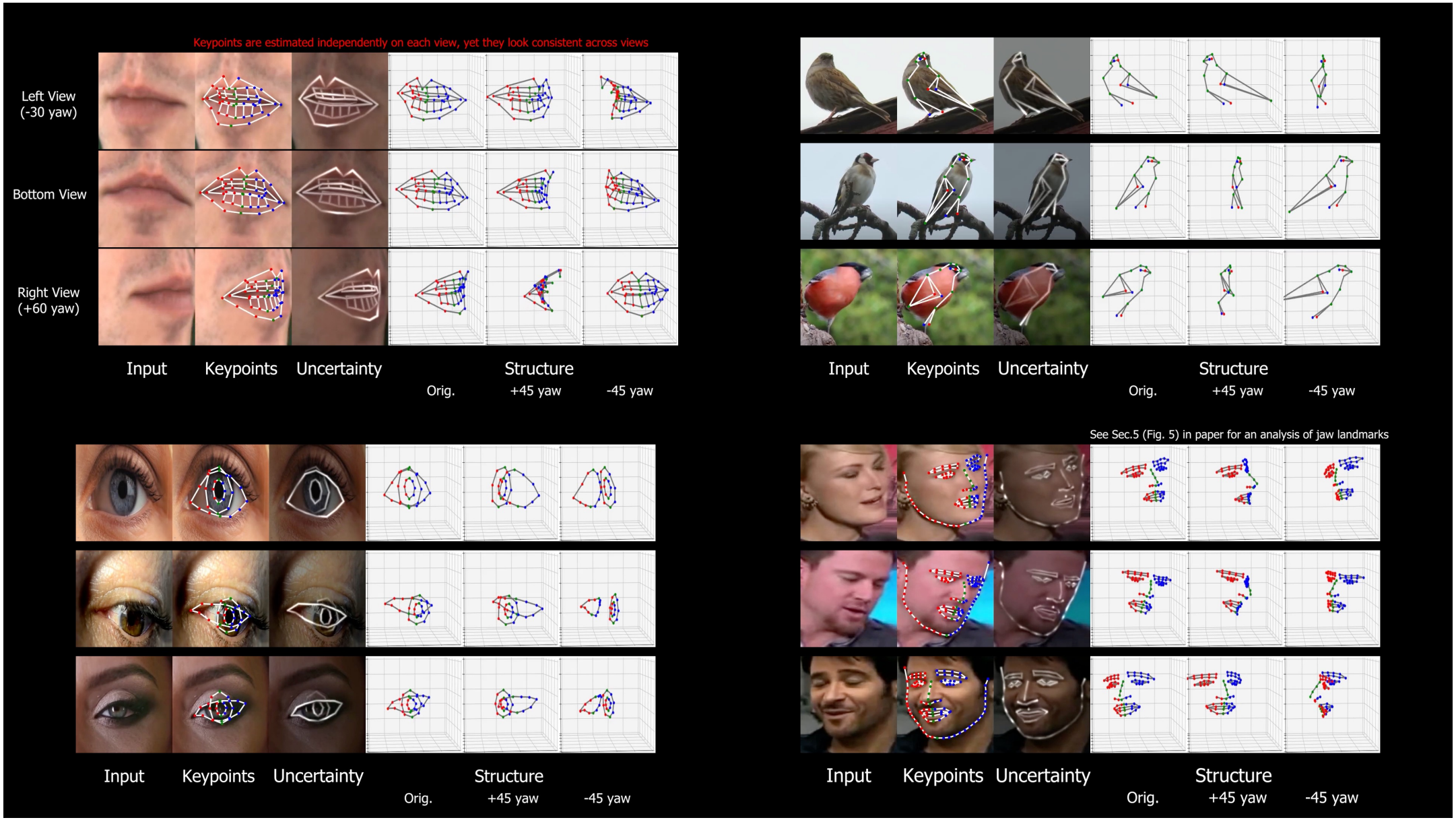
Overview


2D Few-shot Supervision
(Pre-defined Keyframes)
Prediction
GT
1
2D Geometry Constraints
(Equivariance)
Transformed keypoints
Keypoints on transformed image
1
Detector
Input Image
Geometry Aware Keypoints, Uncertainty
Edge Maps
Decoder
Reconstructed Image
Random Masking
Masked Image

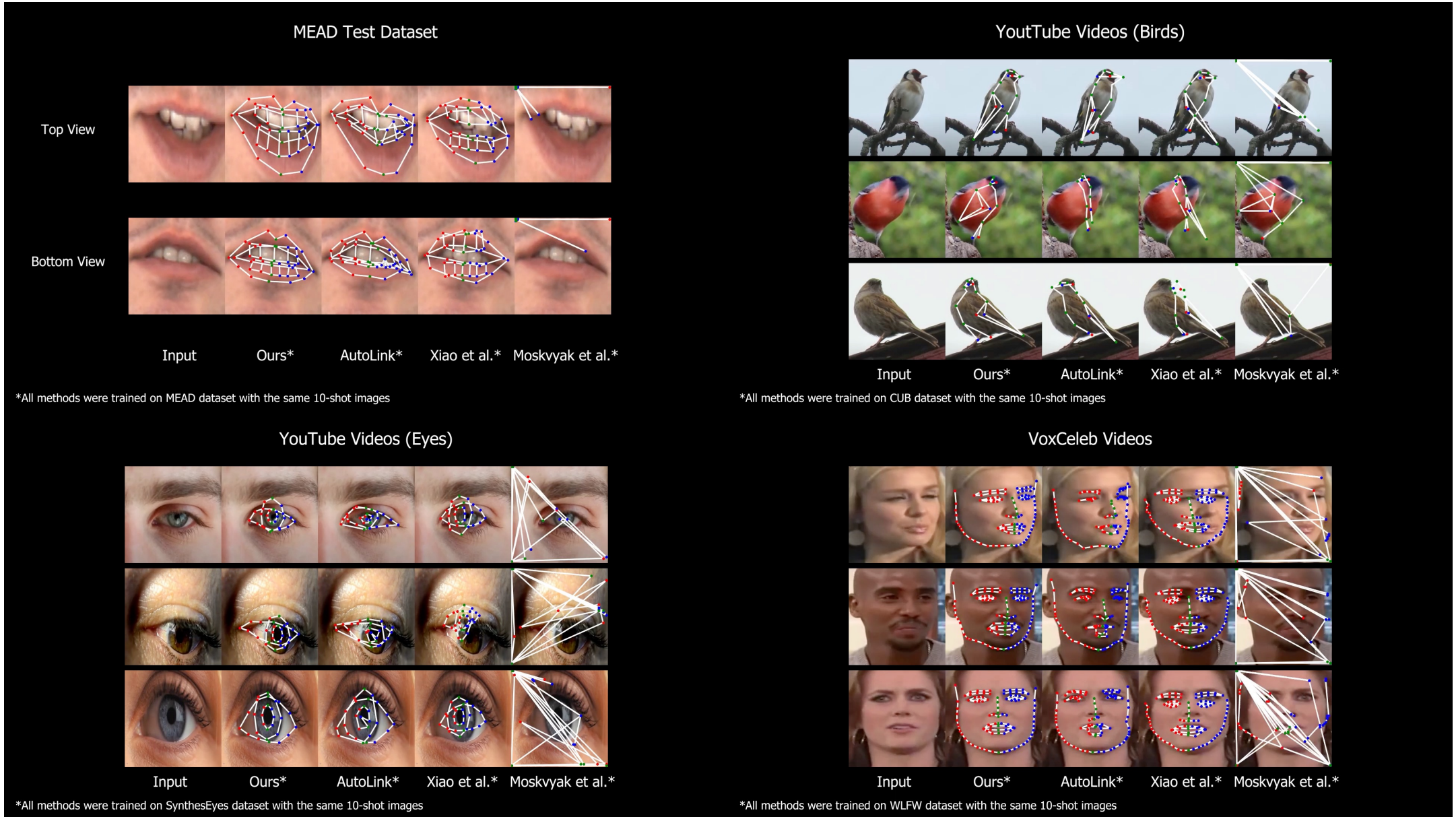
Overview


2D Few-shot Supervision
(Pre-defined Keyframes)
Prediction
GT
1
2D Geometry Constraints
(Equivariance)
Transformed keypoints
Keypoints on transformed image
1
3D Aware Geometry Constraints
(Reprojection Error)
[R | t]
Prediction
Closest Transformed Match in Batch
1
Input Image
Detector
Geometry Aware Keypoints, Uncertainty
Edge Maps
Decoder
Reconstructed Image
Random Masking
Masked Image

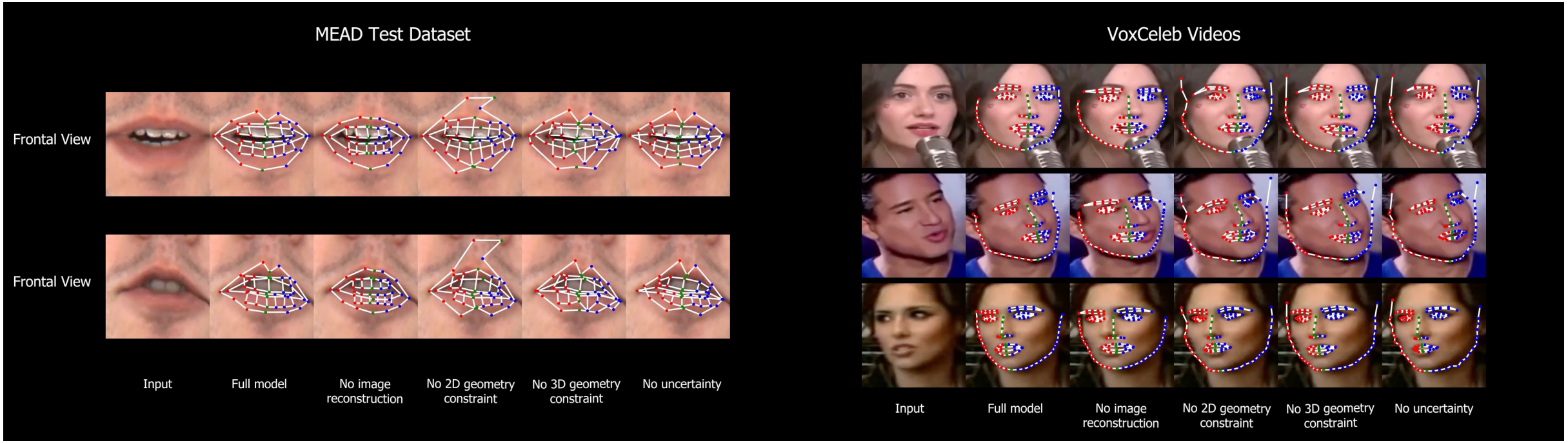
Qualitative Results



Qualitative Comparison



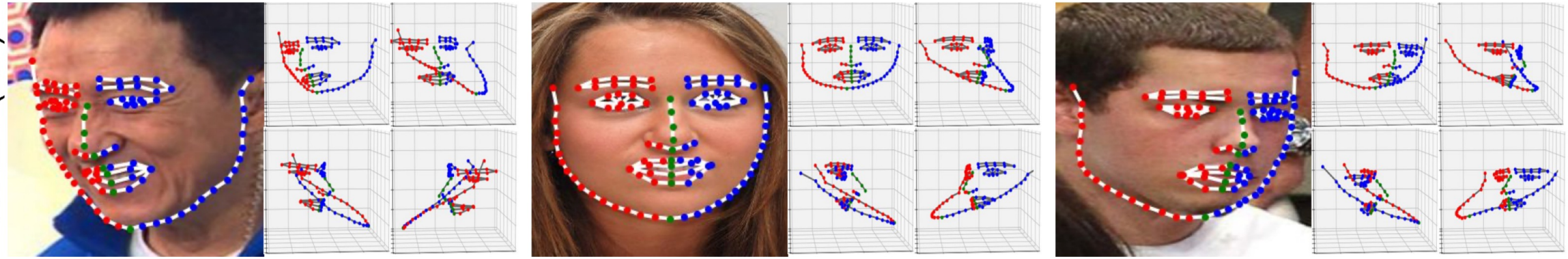
Ablation Tests



- No image reconstruction: overfit to a fixed structure
- No 2D geometry constraint: generating extreme outliers
- No uncertainty / 3D geometry constraint : overfit to visible regions

Ablation Tests

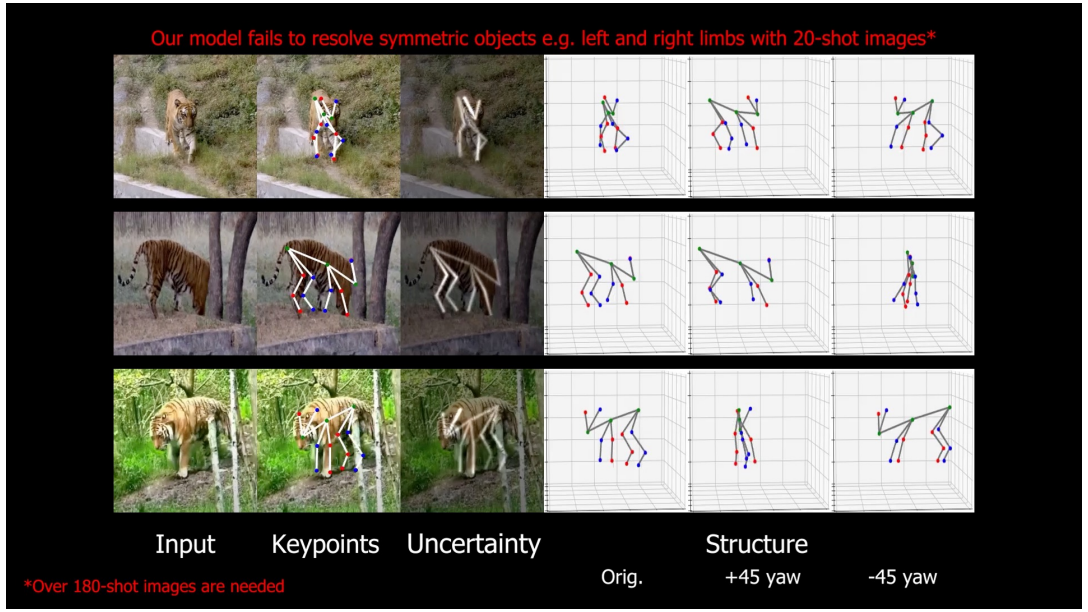
WFLW (2D)



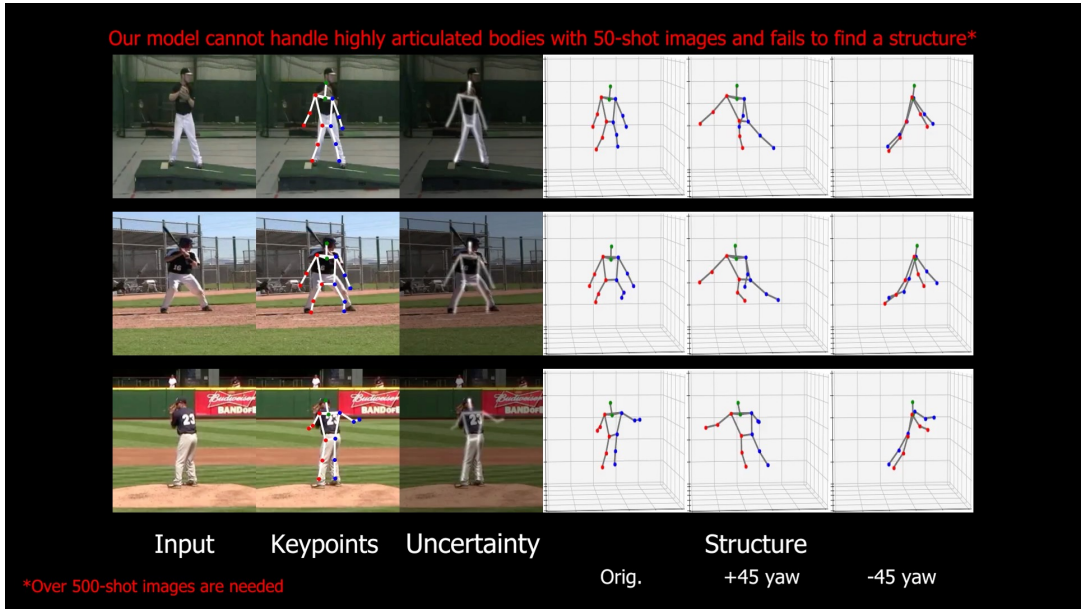
If the annotated keypoints are not consistent in 3D, the predicted 3D keypoints are deformed

Limitations

Fails to resolve symmetric objects



Cannot handle highly articulated bodies



Both problems can be resolved by adding hundreds of labels