WEAK MULTI VIEW SUPERVISION

FOR SURFACE MAPPING ESTIMATION

*__Nishant Rai__[1,2]
Rodrigo Ortiz Cayon[1]

*Aidas Liaudanskas[1]
Matteo Munaro[1]

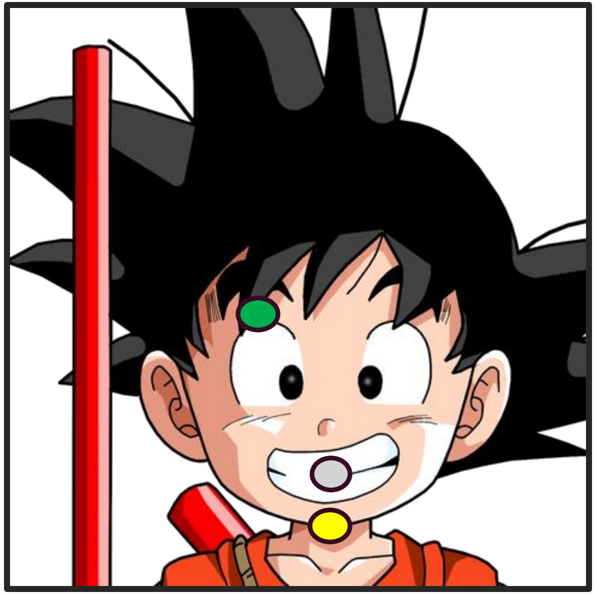
Srinivas Rao[1]
Stefan Holzer[1]

1 FYUSION

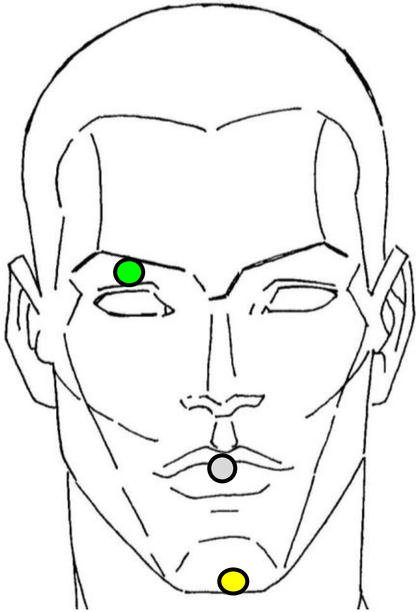
2 Stanford University

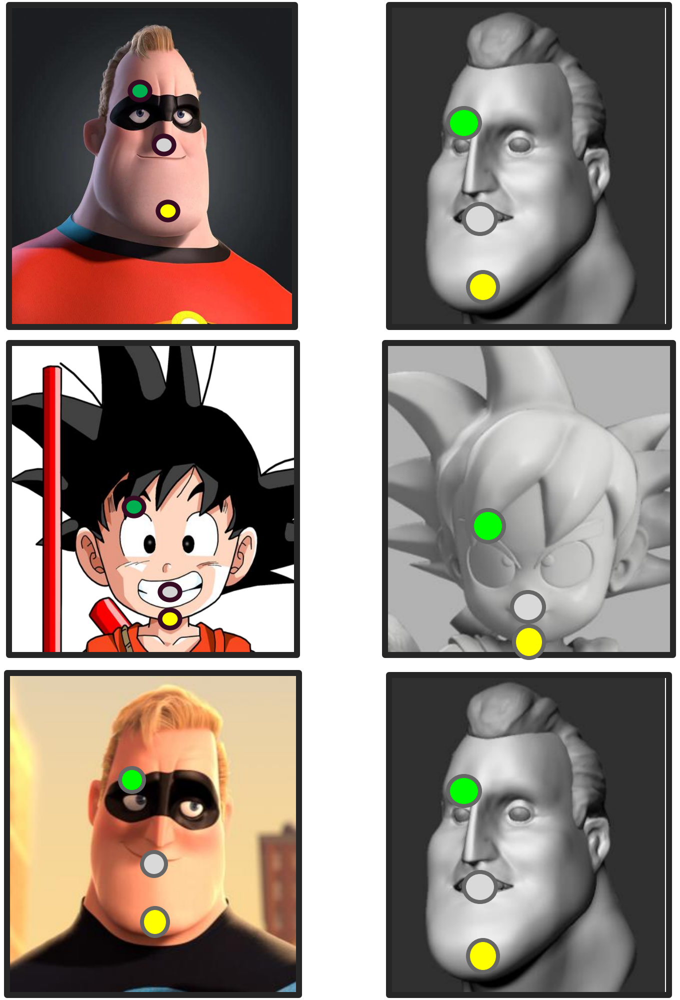
Nishant and Aidas have contributed equally
Work done during Nishant's internship at Fyusion

WHAT WE WANT TO DO?



Dense Correspondence





Middle of Chin

Upper left Eye

Middle of mouth/lips

WHAT WE WANT TO DO?

- Now that we have covered the basics
- The next question is how can a model learn to do this?
- How to predict which pixel maps to which point on our surface?

WHAT WE WANT TO DO?

- Earlier approaches rely heavily on supervision
- Recent work by Kulkarni et al [2] tackles this problem without the need for dense supervision
- Learn models predictive the surface mapping through consistency cycles
- However still limitations related to a fixed template shape

WHAT WE WANT TO DO?

- We propose our approach which simultaneously predicts the underlying 3D shape
- Allows for using multiple views of instances to improve performance
- **Aim:** Learn dense correspondence along with 3D structure under a multi-view setting with minimal supervision

WHY DO WE WANT THIS?

- Currently need too much data to learn!
- Need to create labels for each catergory specific dataset
- Not possible to label all such images with 3d annotations manually!
- Extending to new categories is very effort intensive

DATASET

Synthetic MultiView images with Dense Annotations for evaluation

APPROACH

Weakly supervised learning of surface mapping and shape through Multi View images

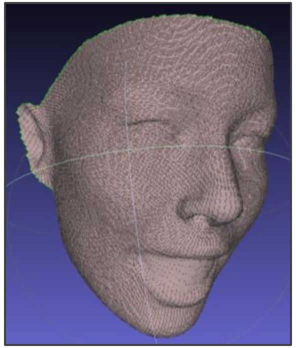
DATASET

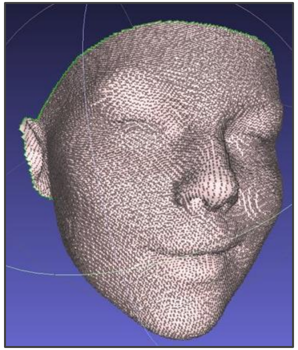
Synthetic MultiView images with Dense Annotations for evaluation

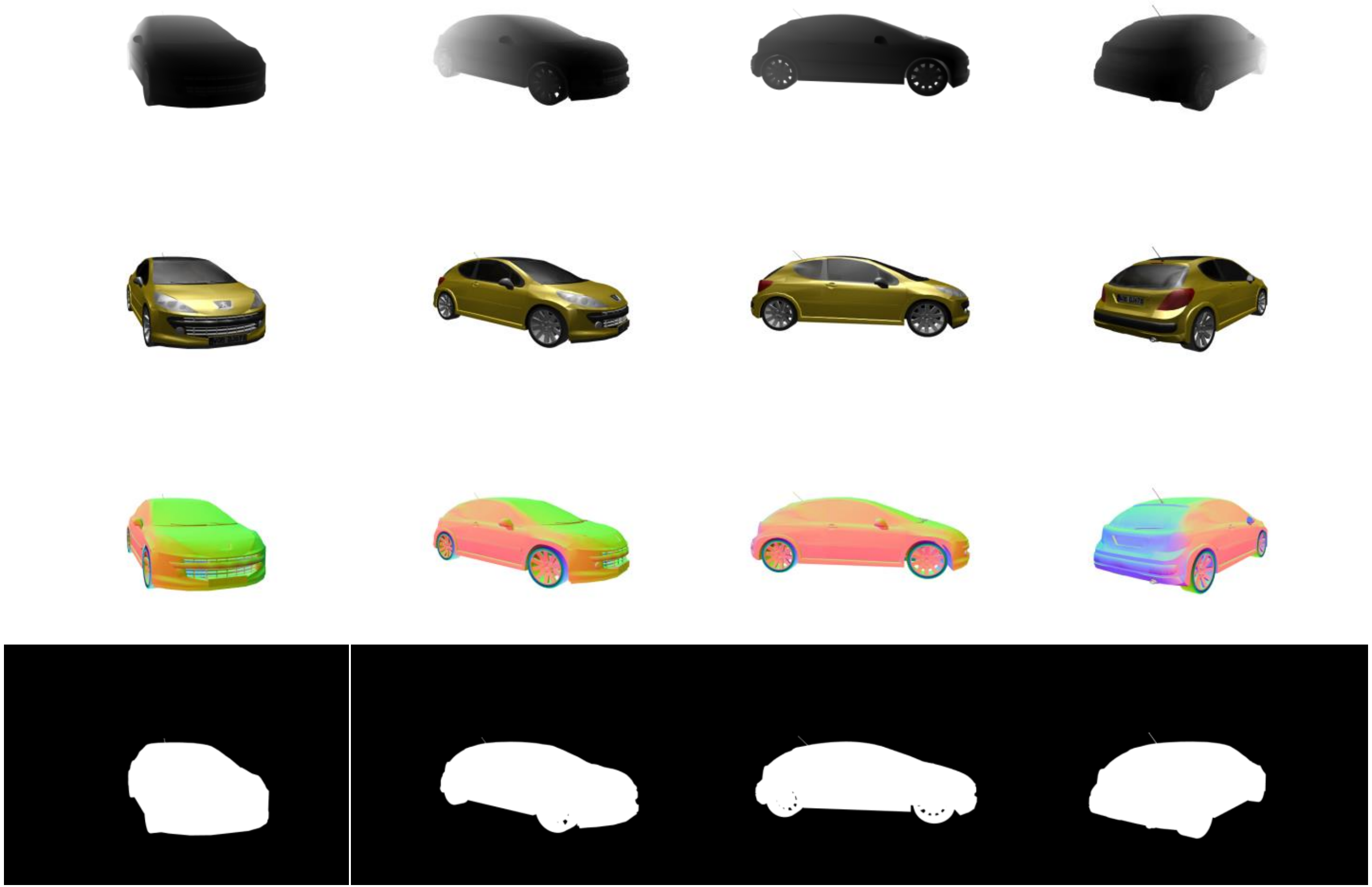
APPROACH

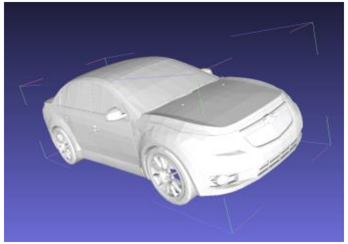
Weakly supervised learning of surface mapping and shape through Multi View images

DATASET

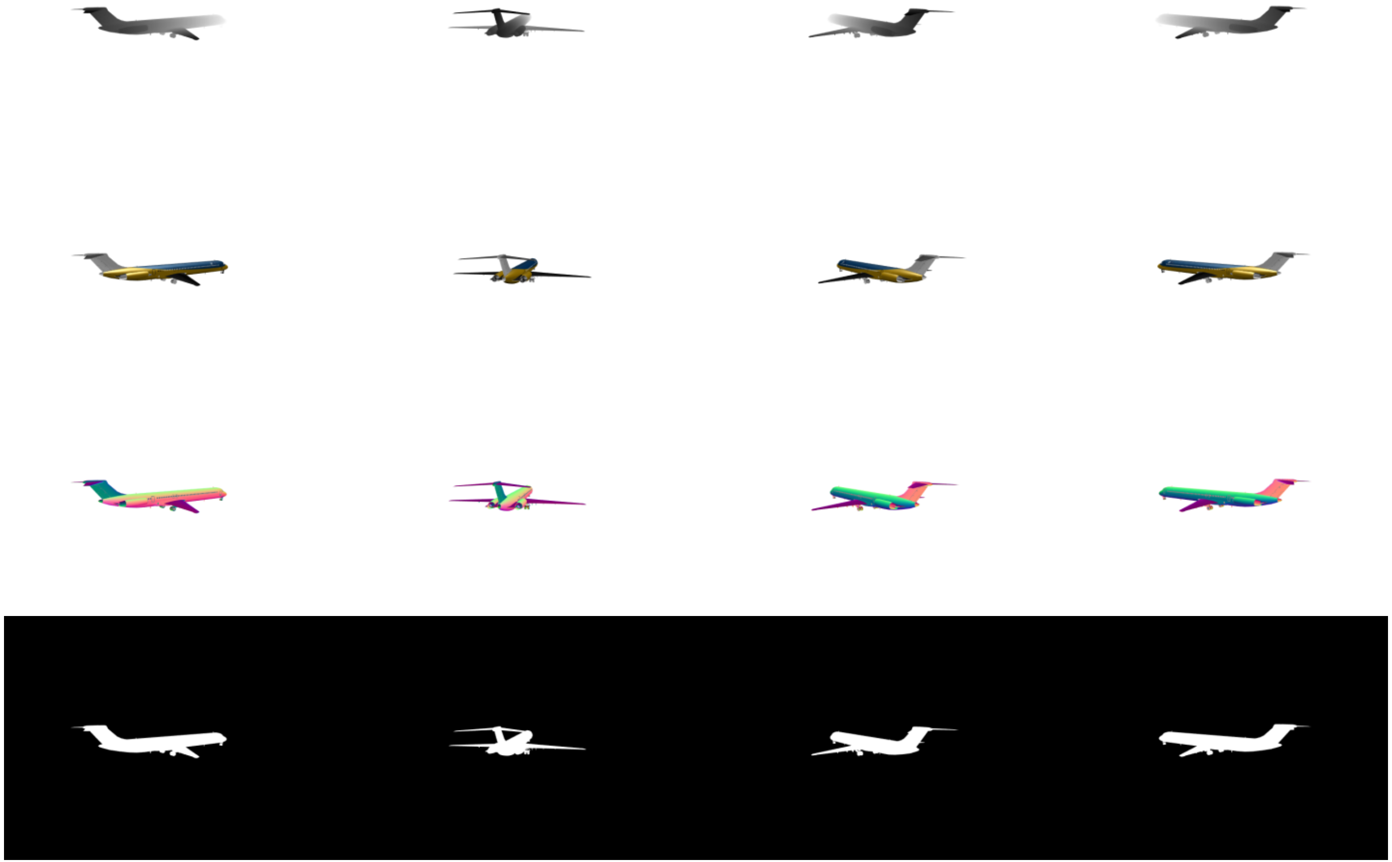








Cars



Planes

DATASET

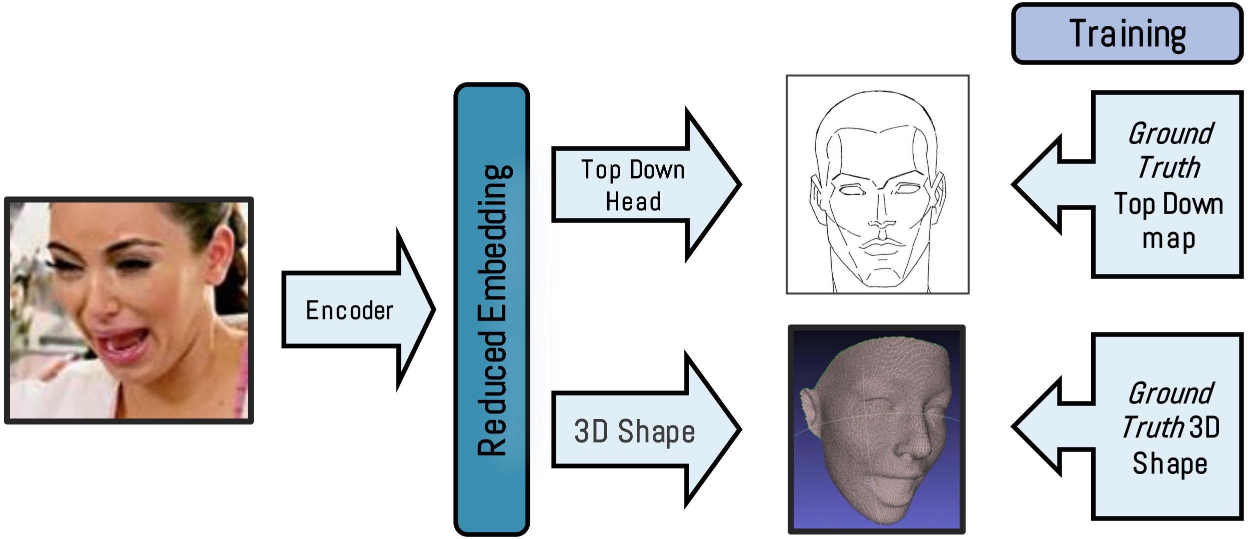
Synthetic Multi View images with Dense Annotations for evaluation

APPROACH

Weakly supervised learning of surface mapping and shape through Multi View images

HOW DO WE DO IT?

- Network with Encoder, UV head (a.k.a. Top-Down) and 3D shape head (a.k.a. PosMap)



HOW DO WE DO IT?

- Both heads trained using ground truth labels
- Give them the expected output for each instance
- Don’t want to use supervision anymore!
- How to predict something without exactly knowing its value?

Reprojection Cycle

Deformed Meshes

HOW DO WE DO IT?

Multi View Consistency

Reprojection Cycle

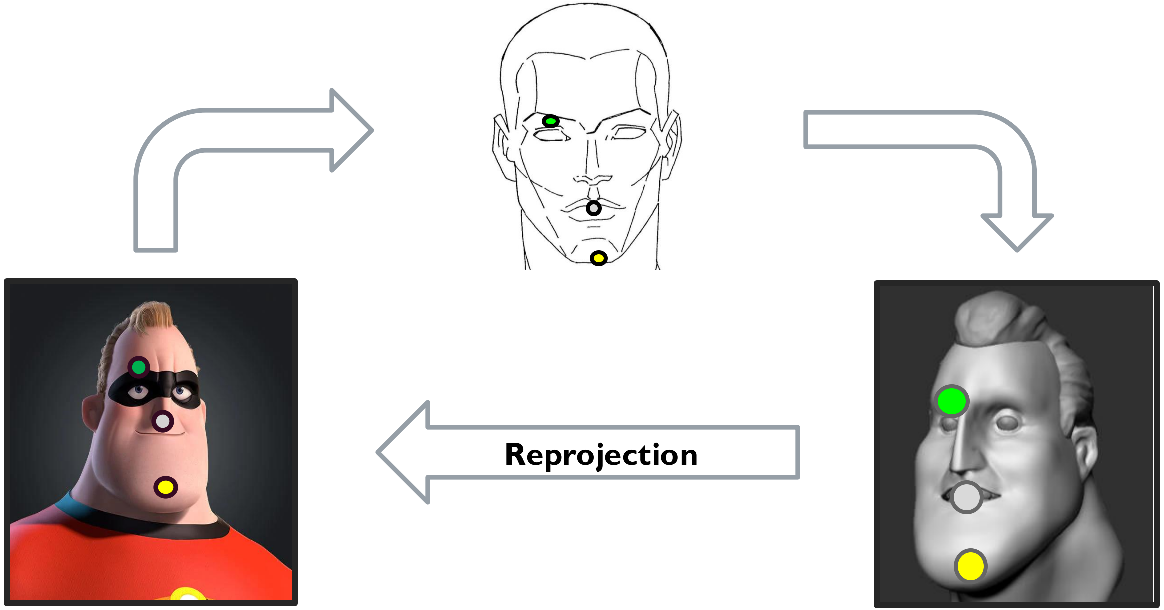
Deformed Meshes

HOW DO WE DO IT?

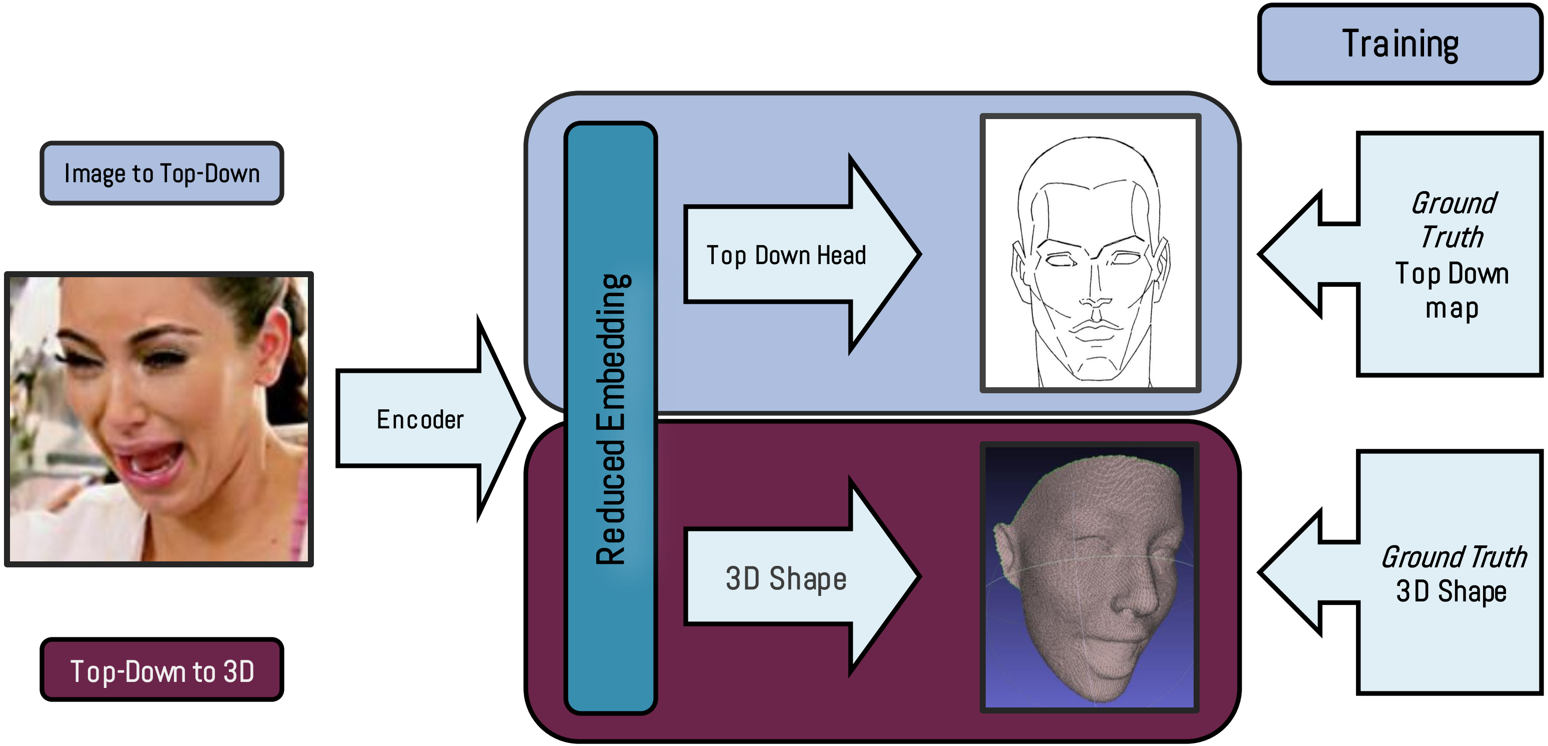
Multi View Consistency

REPROJECTION

- **Assume** we know the true top-down mapping and underlying 3D car shape

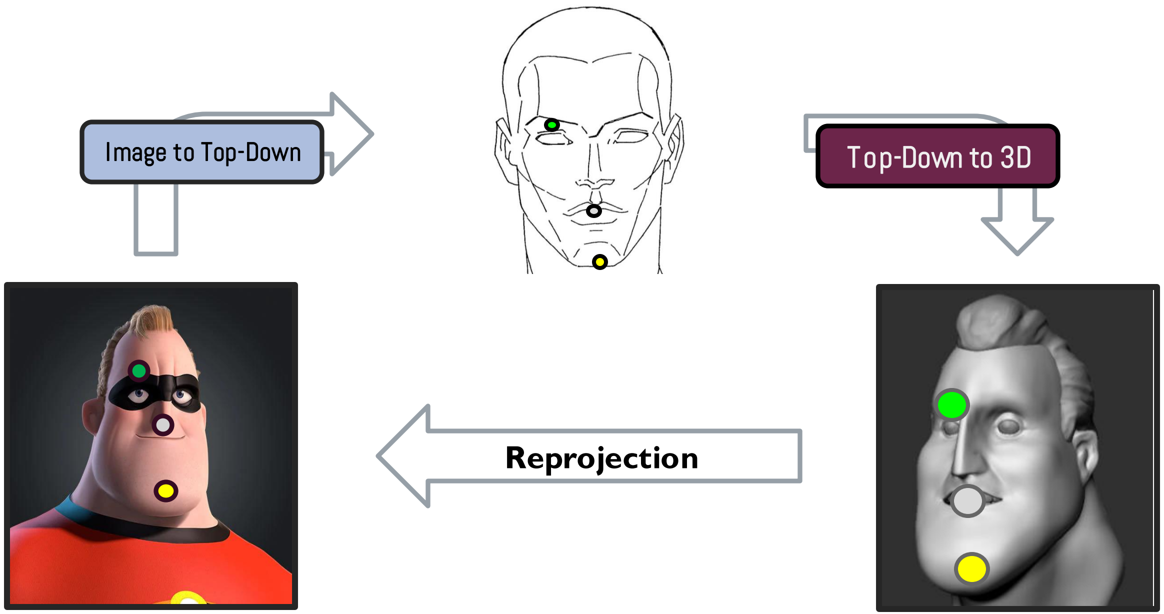


REPROJECTION

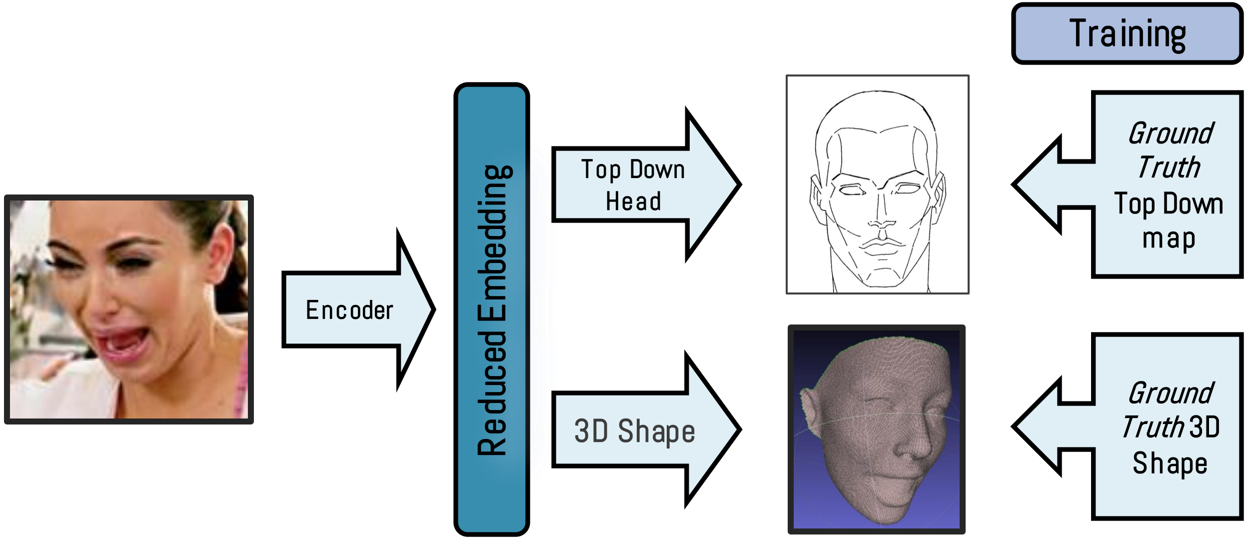


REPROJECTION

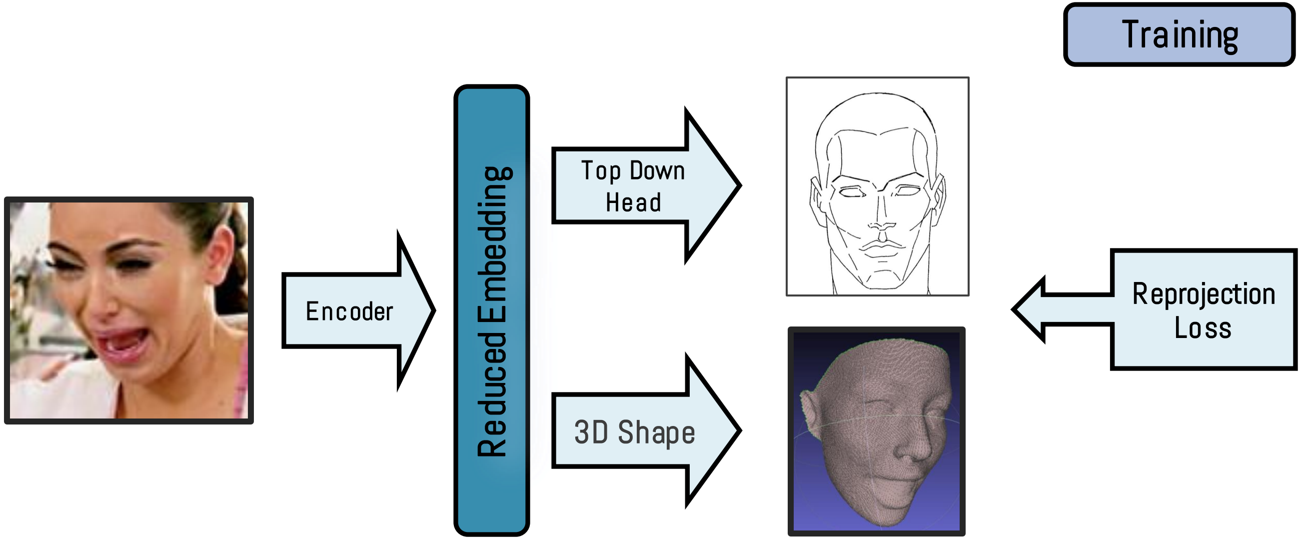
- The network can directly be part of the reprojection cycle!



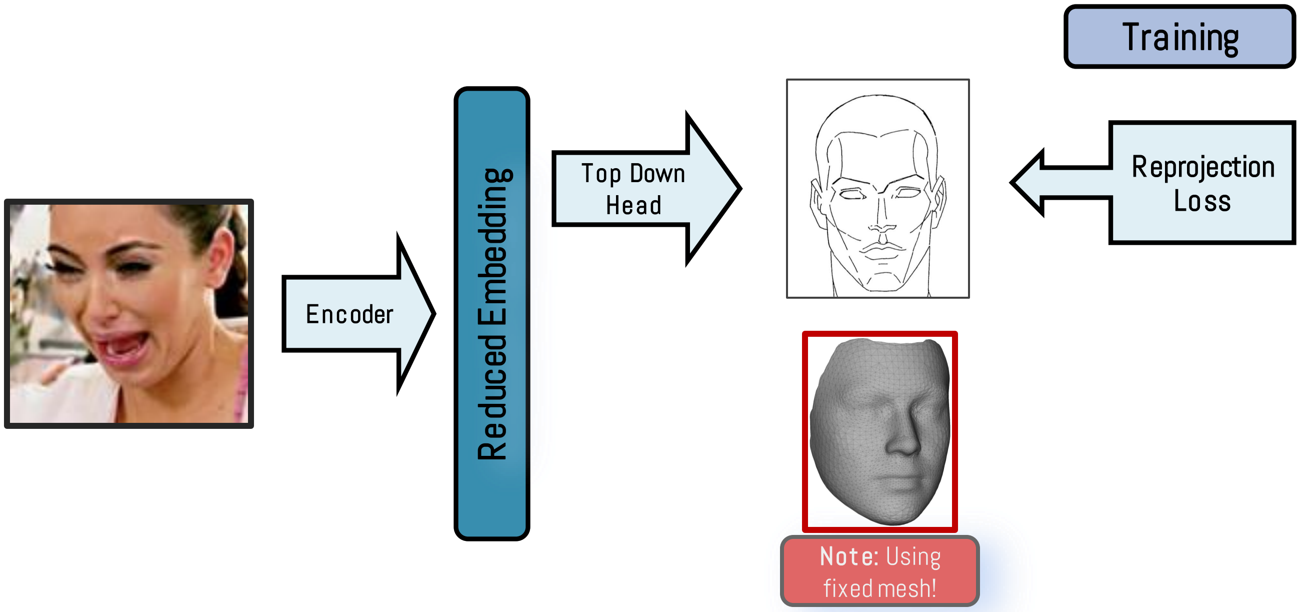
HOW DO WE DO IT - PUTTING IT TOGETHER



HOW DO WE DO IT - PUTTING IT TOGETHER



HOW DO WE DO IT - PUTTING IT TOGETHER



REPROJECTION - RESULTS

Approach	UV-Pck@ 0.01	UV-Pck@ 0.03	UV-Pck@ 0.1	AUC
Learning UVs with fixed mesh	5.3	32.2	90.6	94.0
Learning only UVs	**12.1**	**48.6**	**91.1**	**94.8**
Learning with dense labels	55.1	94.9	99.5	98.7

Benefits of reprojection compared to full supervision

Reprojection Cycle

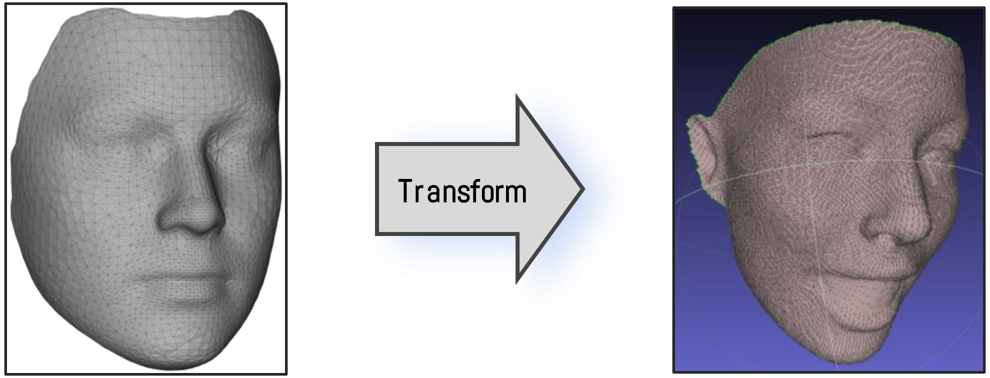
Deformed Meshes

HOW DO WE DO IT?

Multi View Consistency

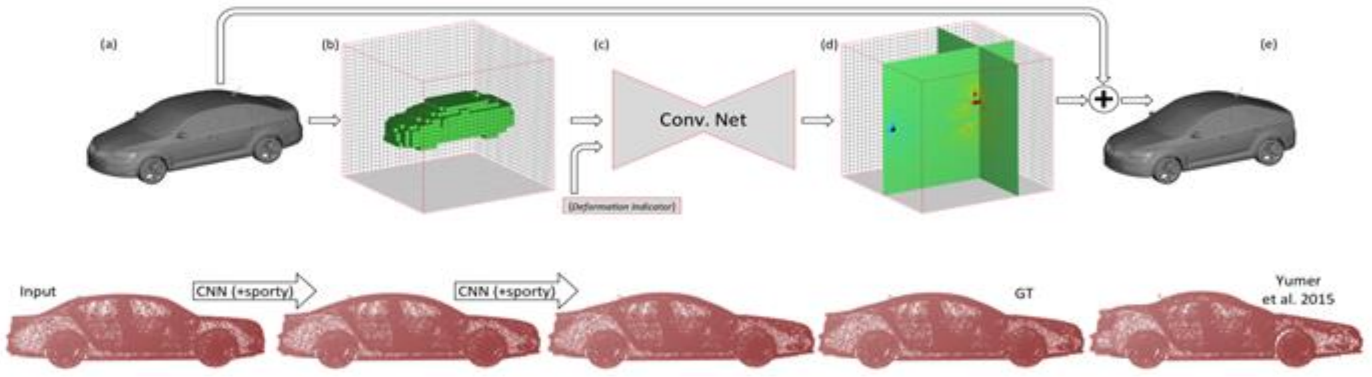
DEFORMED MESHES

- Revisit our fixed mesh assumption
- Is it possible to constrain predicting 3D structures?
- **Solution:** Predict deformation over mean shape



DEFORMED MESHES

- Model deformation as a cubic field
- Predict a 32x32x32 cube representing deformations
- Can deform different regions differently



DEFORMED MESHES - RESULTS

Approach	PosMap-Pck@		
	AUC	0.1	0.03
Mean-Fixed Mesh	96.2	97.6	42.7
Unconstrained	97.6	99.8	74.8
Deformed Mesh	**98.0**	**99.9**	**82.0**

Benefits of our deformation module

Reprojection Cycle

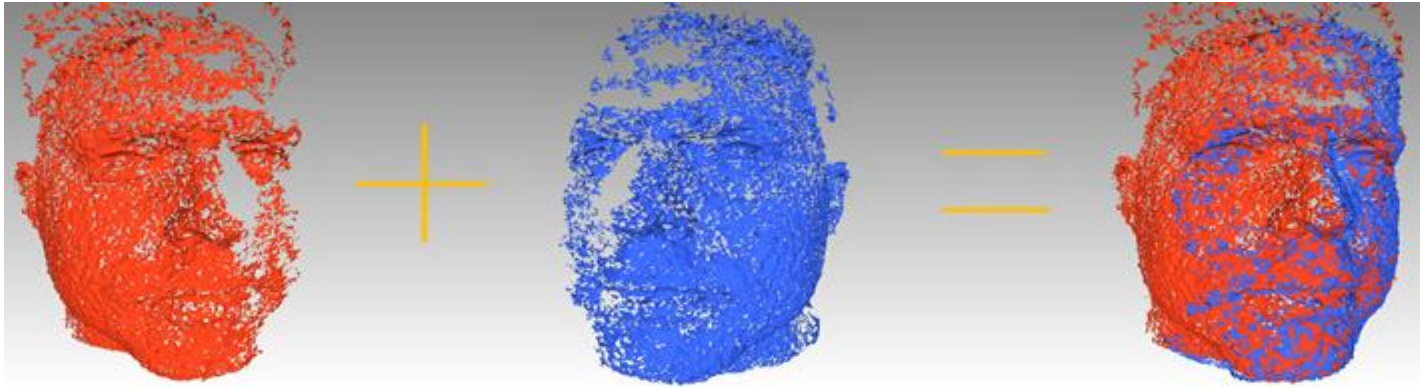
Deformed Meshes

HOW DO WE DO IT?

Multi View Consistency

MULTI VIEW CONSISTENCY

- Haven't explored relationships between different viewpoints of same object yet
- Multi-view contains more information than a single-view image

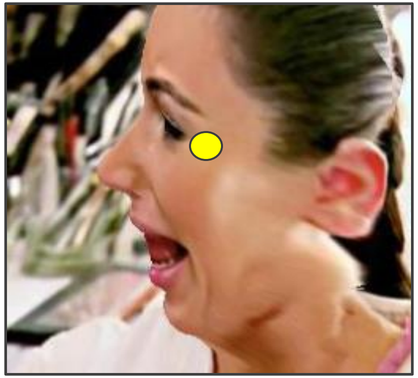


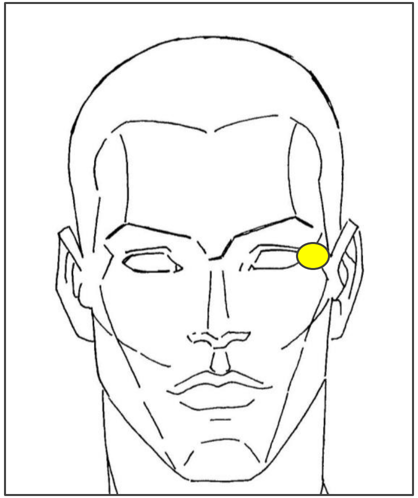
Use **Multi-View** Information to get the **complete** picture

Image taken from [3]

MULTI VIEW CONSISTENCY

- Can we argue something about the same car part in two different images?
- Yes! They should map to the same point in the top-down view





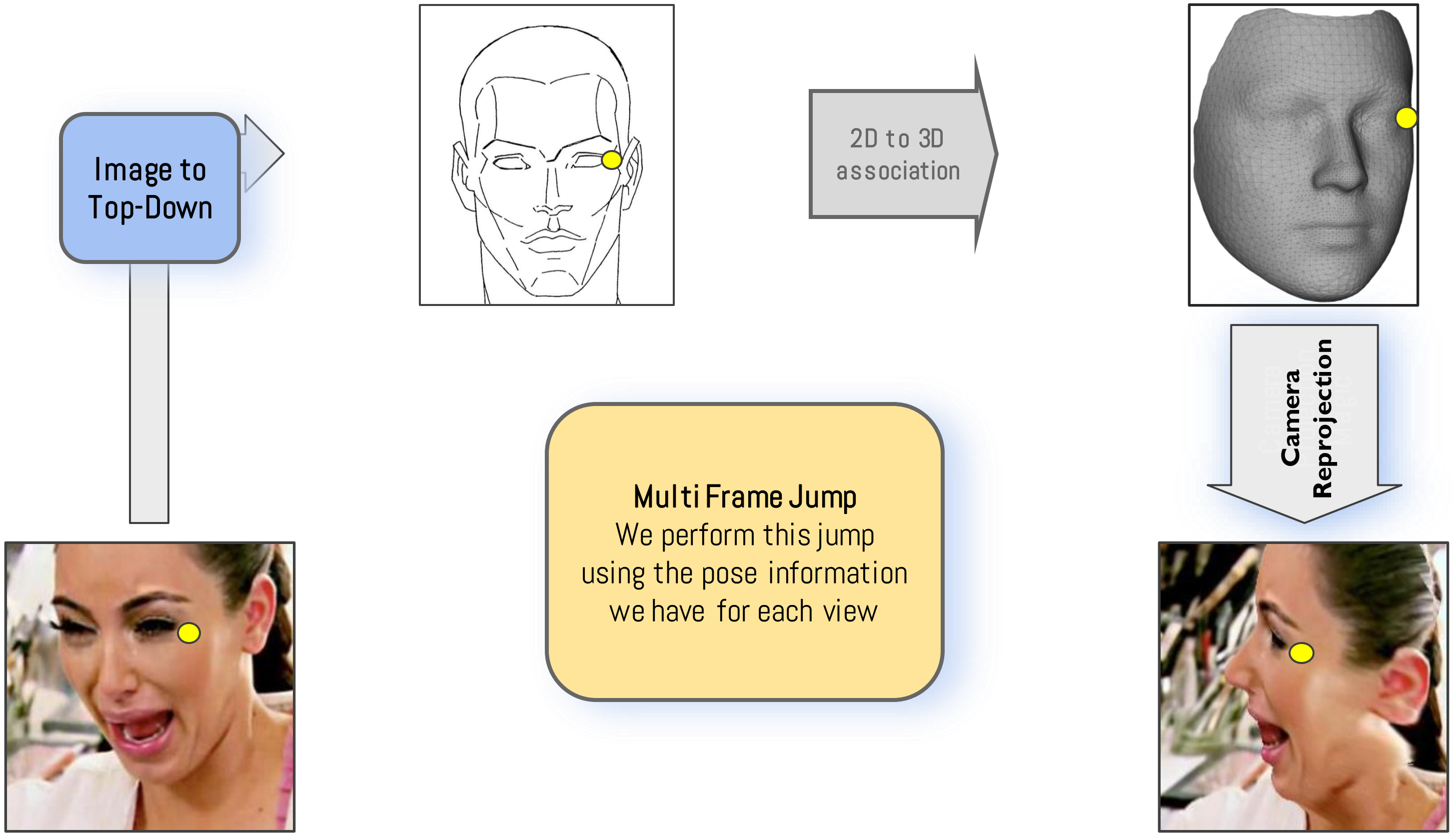


Image to Top-Down
2D to 3D association
Camera Reprojection
Multi Frame Jump
We perform this jump using the pose information we have for each view

MULTI VIEW CONSISTENCY - RESULTS

Approach	UV-Pck@ 0.01	0.03	0.1	AUC
Single-view Reprojection with Fixed Mesh	5.3	32.2	90.6	94.0
Multi-view Reprojection with Fixed Mesh	**5.8**	**34.0**	**90.8**	**94.2**

Benefits of multi view reprojection

Reprojection Cycle

Deformed Meshes

OVERALL MODEL

Multi View Consistency

RESULTS

Approach	UV-Pck@ 0.01	0.03	0.1	AUC
Single-view Reprojection with Fixed Mesh	5.3	32.2	90.6	94.0
Multi-view Reprojection with Fixed Mesh	**5.8**	**34.0**	**90.8**	**94.2**
Deformed Single-view Reprojection	**13.5**	57.8	96.0	95.7
Deformed Multi-view Reprojection	13.4	**58.4**	**96.2**	**95.9**

RESULTS

Approach	UV-Pck@ 0.03	UV-Pck@ 0.1	PosMap-Pck@ 0.03	PosMap-Pck@ 0.1
CSM*	32.2	90.6	71.7	**98.6**
Our Approach	**58.4**	**96.2**	**72.7**	**98.6**
Fully Supervised	94.9	99.5	82.0	99.8

CONCLUSIONS AND FUTURE WORK

- Learned how to exploit reprojection cues in a multi view setting
- Able to effectively learn 3D structure as well as see improvements in surface mapping
- We hope our released dataset encourages research in this direction
- Allows for stronger evaluations of comparable methods

REFERENCES

1. Deformation Fields: http://www.meyumer.com/deformation-flow-3D-conv-net.html
2. Canonical Surface Mapping: https://nileshkulkarni.github.io/csm/
3. Identity recognition using 4D Facial Dynamics: dynface4d.isr.uc.pt/database.php
4. What is Camera Calibration: http://www.mathworks.com/help/vision/ug/camera-calibration.html

THANK YOU!
QUESTIONS?