



# Improving self-supervised 3D face reconstruction with few-shot transfer learning

Martin Dornier, Philippe-Henri Gosselin, Yann Ricquebourg, Christian Raymond, Bertrand B. Coüasnon

## ► To cite this version:

Martin Dornier, Philippe-Henri Gosselin, Yann Ricquebourg, Christian Raymond, Bertrand B. Coüasnon. Improving self-supervised 3D face reconstruction with few-shot transfer learning. ACM MIG 2023, Nov 2023, Rennes, France. The 16th Annual ACM SIGGRAPH Conference on Motion, Interaction and Games. hal-04290214

HAL Id: hal-04290214  
<https://hal.science/hal-04290214v1>

Submitted on 16 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## 1. INTRODUCTION

### 1.1 3D face reconstruction

**Principle:** Get a face 3D mesh from a monocular face image



### 1.2 Problem: lack of annotated data

- Supervised methods:**

- Need annotated data to train
- 3D annotations are hard to obtain**

- Self-supervised methods:**

- Do not need annotated data
- But tend to predict wrong head pose and scale**

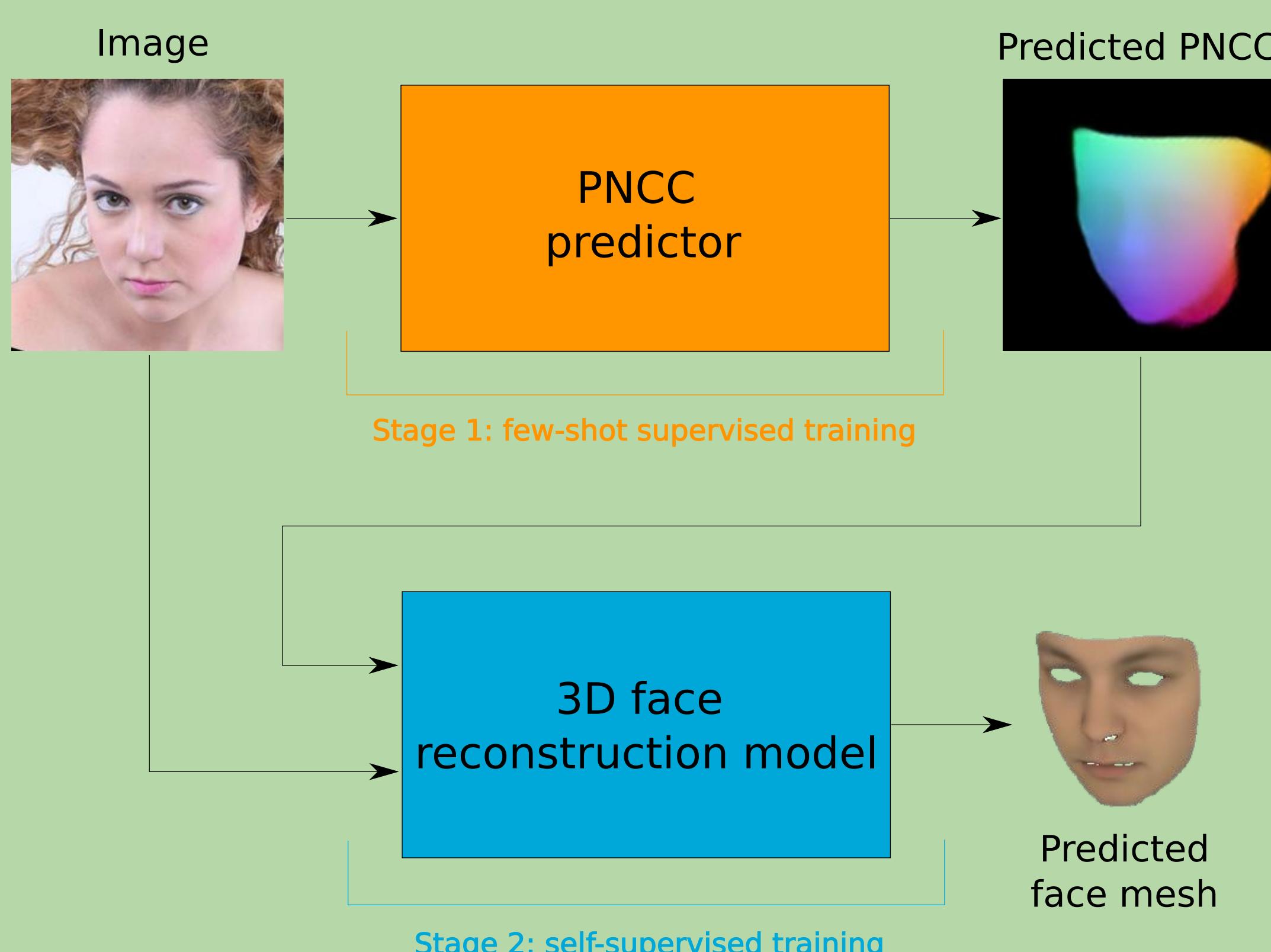
### 1.3 Goal: Improve self-supervised methods predicted head pose and scale

- Add **supervised 3D** information
- But use **as few as possible** annotated 3D samples

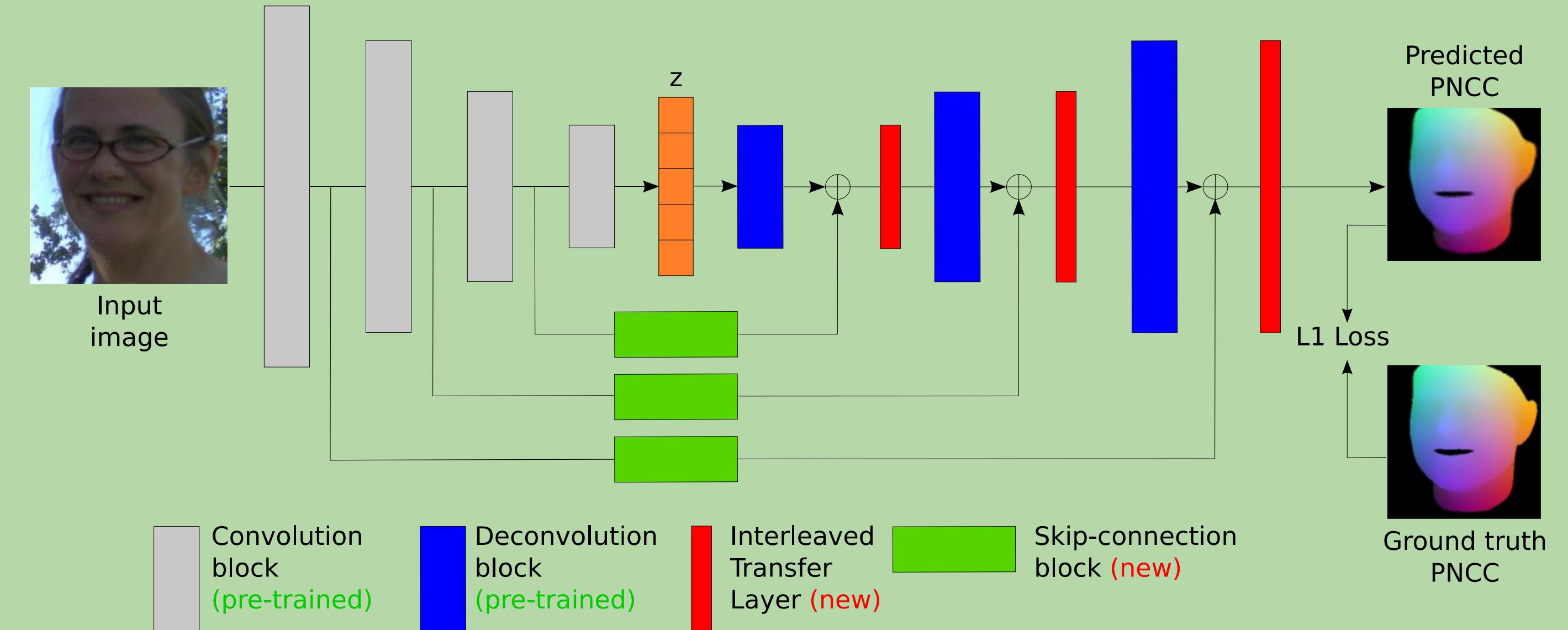
## 2. Method

### 2.1 Principle

- Use the **Projected Normalized Coordinate Code (PNCC)** [1] as supervised information
- 2-stage** pipeline
  - Stage 1:** Train a PNCC predictor with **a few annotated 3D samples**
  - Stage 2:** Add the **PNCC as additional input** to a self-supervised 3D face reconstruction model



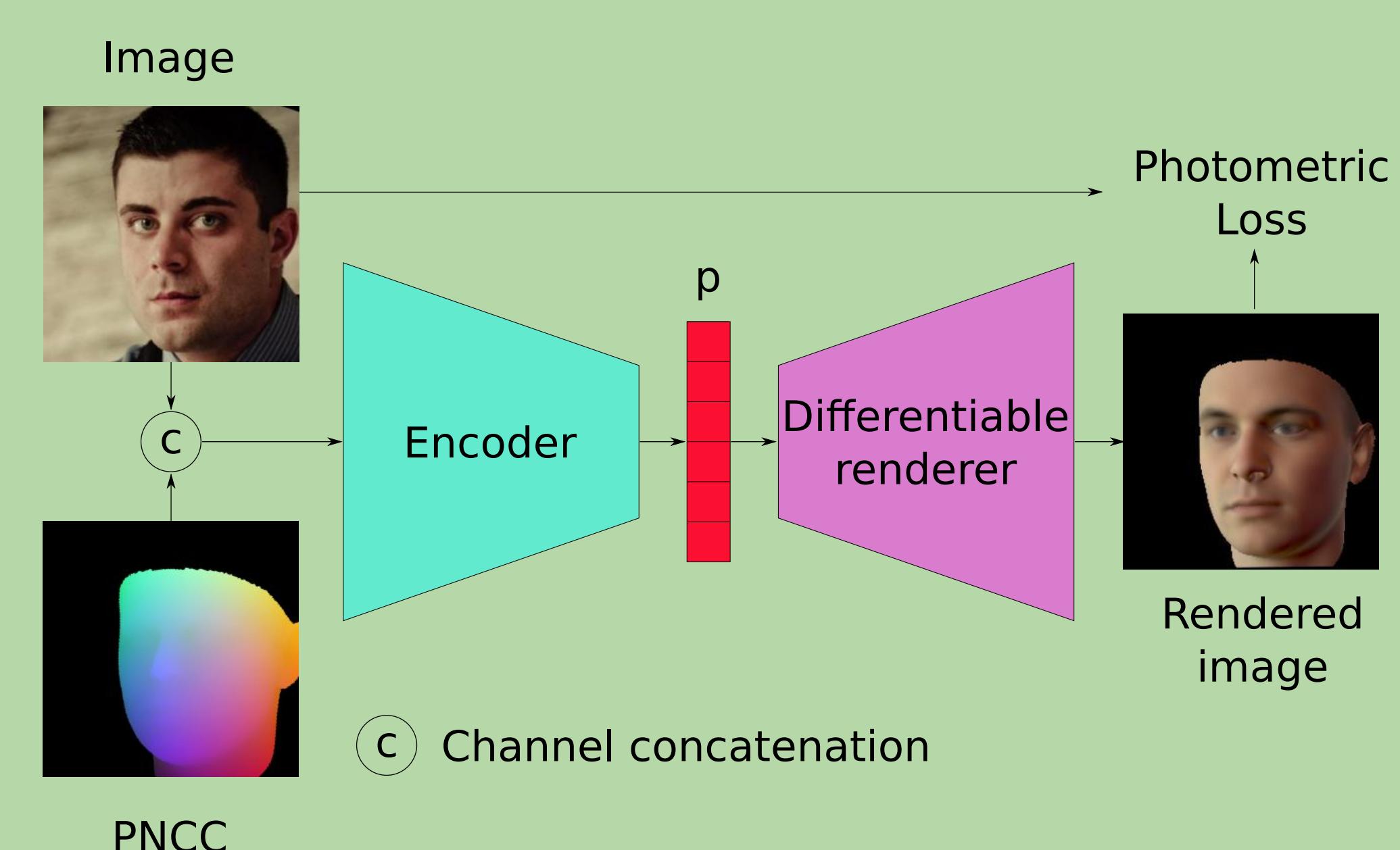
### 2.2 PNCC predictor



- Architecture based on SCAF [4]

- Pre-trained face autoencoder** adapted to the PNCC prediction task
- Interleaved Transfer Layers (ITLs)** [3] and **skip-connections** are added for the supervised PNCC prediction task
- ITLs adapt the decoder flow to the PNCC prediction task
- Skip-connections improve the quality of the generated PNCCs
- Can be trained with **limited annotated data**

### 2.3 Enhanced 3D face reconstruction model



- 3D face reconstruction model based on the **self-supervised model MoFa** [2]
- The encoder predicts the 3D face parameter vector (3D face reconstruction goal)
- The decoder renders the face from the parameter vector
  - Only used for training
- We enhance MoFa [2] with an **additional PNCC input** to help it **predict better head pose and scale**
- The **face image** is **concatenated** with the **PNCC** predicted with our PNCC predictor

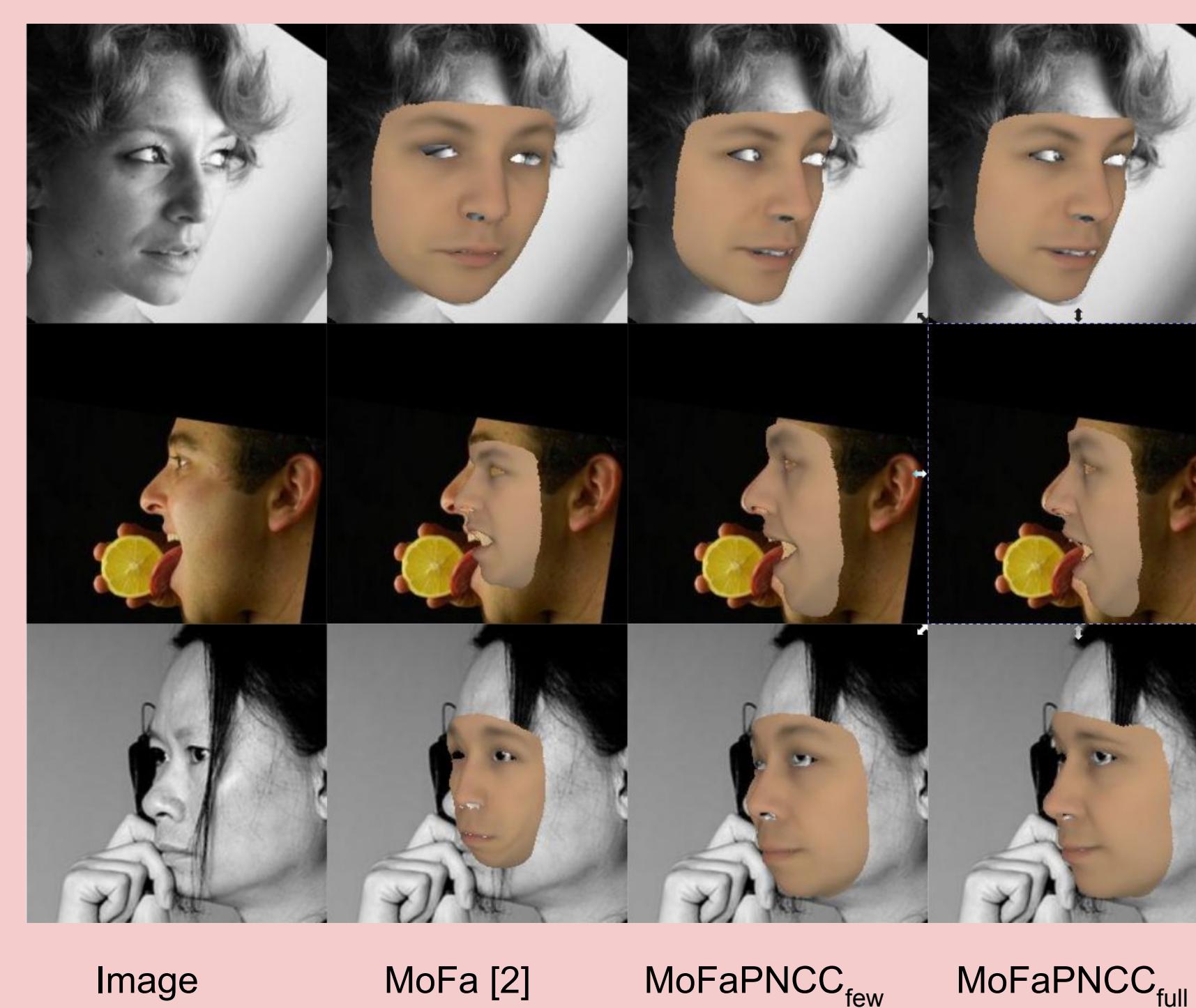
## 3. Results

### 3.1 Quantitative results

Method	Dense 2D ↓	Dense 3D ↓	Yaw MAE ↓
MoFa [2]	4.31	5.85	4.97
MoFaPNCC <sub>few</sub> (Ours)	4.20	5.66	4.95
MoFaPNCC <sub>full</sub> (Ours)	4.12	5.48	4.66

- MoFaPNCC<sub>few</sub> uses PNCC predictions from our PNCC predictor trained with **only 50 annotated samples** of 300W-LP [1] dataset
- MoFaPNCC<sub>full</sub> uses PNCC predictions from our PNCC predictor trained on the whole 300W-LP
- Both models **improve over MoFa** [2] for the **dense alignment** (Dense 2D and 3D) and **head pose** (Yaw MAE) error metrics

### 3.1 Qualitative results



- Our models predict better head pose and face scale

## 4. Conclusion

- Proposed method for 3D face reconstruction
  - Goal:** improve the predicted **head pose** and **face scale** of self-supervised methods
  - Context:** use **limited annotated data** (50 examples)
  - Principle:** use the **PNCC** [1] to help the 3D face reconstruction model
- Implementation: 2-stage pipeline**
  - Stage 1: PNCC predictor** training
    - Based on a **pre-trained generative model**
    - Adapted to the PNCC task via the addition of **ITLs** and **skip-connections**.
    - Trained with **limited annotated data**
  - Stage 2: Self-supervised 3D face reconstruction model** training
    - Predicted PNCCs as additional input**
    - Results: improved predicted head pose and face scale** compared to baseline

Martin Dornier<sup>1,2</sup>

Philippe-Henri Gosselin<sup>1</sup>

Yann Ricquebourg<sup>2</sup>

Christian Raymond<sup>2</sup>

Bertrand Coüasnon<sup>2</sup>

<sup>1</sup>InterDigital

<sup>2</sup>Univ Rennes, CNRS, IRISA, France

### Contact

martin.dornier@irisa.fr

### References

- [1] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Haolin Shi, and Stan Z Li. 2016. Face alignment across large poses: A 3d solution. In Proceedings of the IEEE conference on computer vision and pattern recognition. 146–155
- [2] Ayush Tewari, Michael Zollhofer, Hyeyoung Kim, Pablo Garrido, Florian Bernard, Patrick Perez, and Christian Theobalt. 2017. MoFa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In Proceedings of the IEEE International Conference on Computer Vision Workshops. 1274–1283
- [3] Björn Browatzki and Christian Wallraven. « 3fabrec: Fast few-shot face alignment by reconstruction ». In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6110–6120
- [4] Martin Dornier, Philippe-Henri Gosselin, Christian Raymond, Yann Ricquebourg, and Bertrand Coüasnon. « SCAF: Skip-Connections in Auto-encoder for Face Alignment with Few Annotated Data ». In: International Conference on Image Analysis and Processing, Springer, 2022, pp. 425–437.