3DV 2018
International Conference on 3DVision
Verona, Italy
September 5 - 8, 2018

Learning to Generate Facial Depth Maps

Stefano Pini, Filippo Grazioli, Guido Borghi, Roberto Vezzani and Rita Cucchiara
{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

softech-ict
Centro Interdipartimentale di Ricerca
Softech: ICT per le Imprese

Facial Depth Maps Generation

*Is it possible to **generate facial depth maps** from the corresponding RGB ones?*

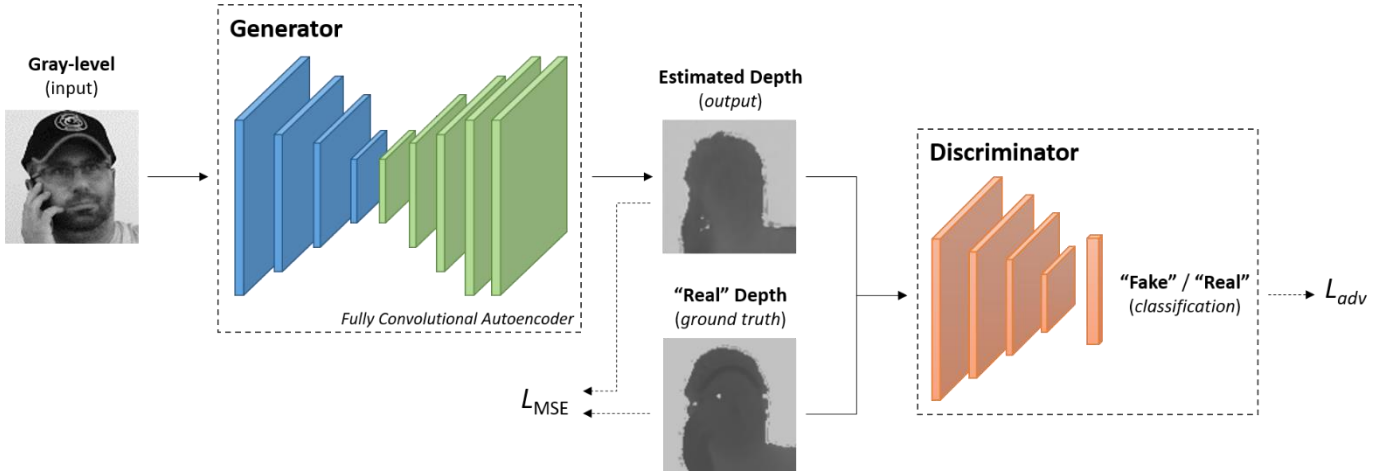
This is one of the **first attempts** to tackle this task

- through a **Conditional GAN**
- **directly** from depth maps
  - no **camera calibration**
  - **no facial landmarks**
  - **no geometric** computation, no **3D models**
  - …

Proposed Method

By following an *image-to-image* approach, we combine the advantages of:

- **Supervised** learning
- **Adversarial** training

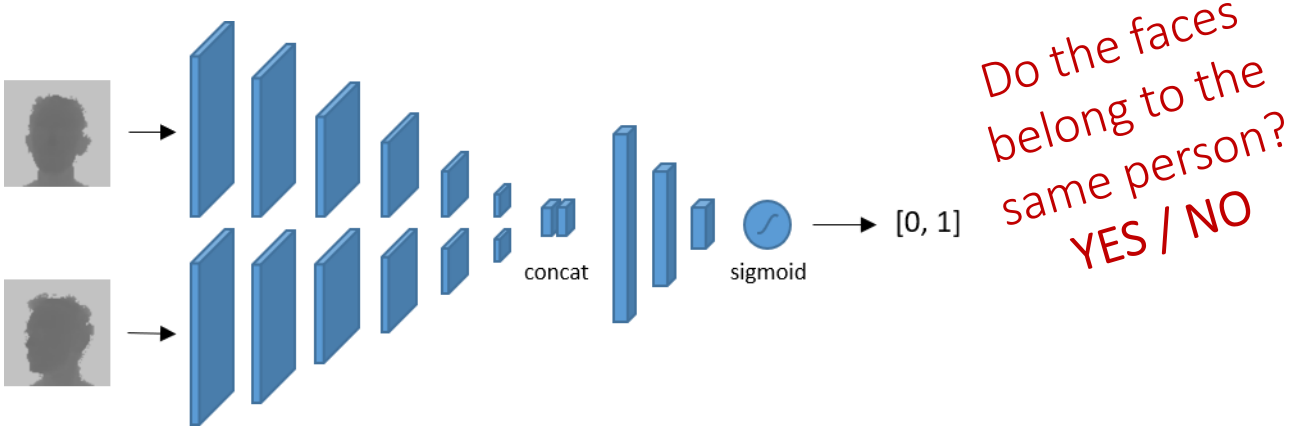


We propose a **Conditional Generative Adversarial Network** that effectively learns to translate intensity face images into the corresponding depth maps.

Metrics

Moreover, we investigate **how to effectively measure the performance** of the system. In particular, we introduce:

- a variety of **pixel-wise metrics**
- a **Face Verification** test
  - a *Siamese* network [1] **trained on the original face depth images**
  - we check if the generated images **maintain the facial distinctive features** of the original subjects



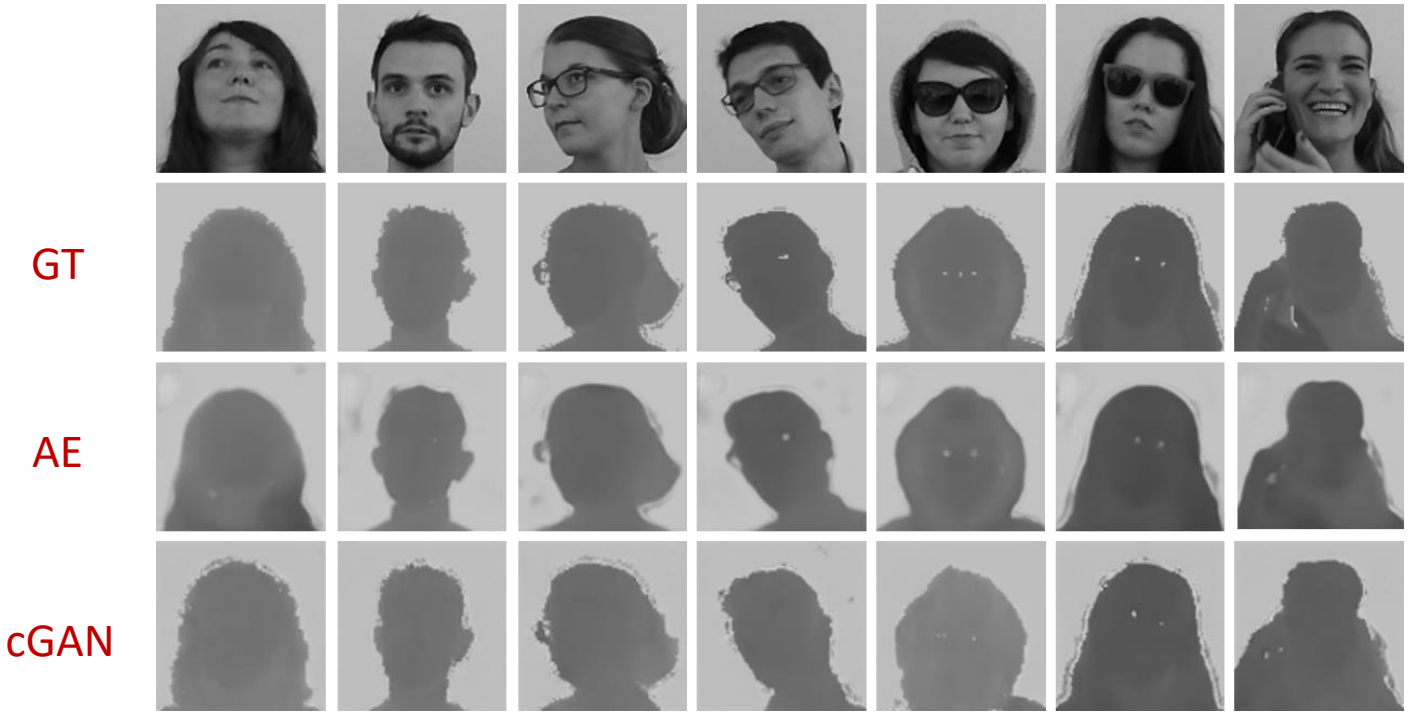
[1] G. Borghi, S. Pini, F. Grazioli, R. Vezzani, R. Cucchiara "Face Verification from Depth using Privileged Information" (BMVC 2018)

Datasets

Two public datasets:

- **Biwi Kinect Head Pose Database** (Microsoft Kinect v1)
- **Pandora** dataset (Microsoft Kinect v2)

are exploited to demonstrate that the proposed model generates **high-quality synthetic depth images**, both in terms of **visual appearance** and **informative content**.



AImage Lab

Results

Pixel-wise metrics

Metrics		Pandora *cGAN*	Pandora *AE*	Pandora pix2pix	Pandora Cui et al.	Biwi *cGAN*	Biwi *AE*	Biwi pix2pix	Biwi Cui et al.
L_1 Norm	↓	**11.792**	16.185	18.172	19.046	10.503	**10.444**	47.191	16.507
L_2 Norm	↓	**1,678.2**	2,224.8	3,109.0	2,093.3	2,368.5	2,342.5	6,661.3	**2,319.8**
Absolute Diff	↓	**0.1019**	0.1441	0.1512	0.1465	**0.1838**	0.1936	0.9062	0.2836
Squared Diff	↓	**2.9974**	5.3891	8.6444	3.9084	**8.7122**	9.0332	100.89	9.3032
RMSE_{lin}	↓	**18.677**	25.213	33.526	22.599	24.865	24.699	72.084	**24.521**
RMSE_{log}	↓	**0.1744**	0.2752	1.0864	0.2105	**0.2932**	0.2970	1.2240	0.3390
$\text{RMSE}_{\text{scale-inv}}$	↓	0.1345	0.2018	1.0774	**0.1301**	0.2687	**0.2642**	1.1759	0.2867
$\delta < 1.25$	↑	**0.8529**	0.6854	0.7802	0.7556	**0.7393**	0.7230	0.4149	0.6395
$\delta < 1.25^2$	↑	**0.9642**	0.8728	0.8978	0.9554	**0.9224**	0.9064	0.5298	0.7943
$\delta < 1.25^3$	↑	0.9915	0.9651	0.9638	**0.9935**	**0.9609**	0.9557	0.6360	0.9311
$\delta < 1.25^{\frac{1}{4}}$	↑	**0.5619**	0.4763	0.5053	0.0740	0.5293	**0.5532**	0.3128	0.3580
$\delta < 1.25^{\frac{1}{3}}$	↑	**0.6113**	0.5047	0.5615	0.1036	0.5727	**0.6034**	0.3274	0.4549
$\delta < 1.25^{\frac{1}{2}}$	↑	**0.6954**	0.5547	0.6496	0.5509	0.6348	**0.6606**	0.3513	0.5189
Face Verification	↑	**0.7247**	0.6570	0.5315	0.6442	**0.6251**	0.6043	0.5422	0.5966

Face Verification Accuracy

	$\{S_i\}_{i=1,2,3}$ *original*	$\{S_i\}_{i=1,2,3}$ *generated*	$\{S_i\}_{i=4,5}$ *original*	$\{S_i\}_{i=4,5}$ *generated*	$\{S_i\}_{i=1,2,3,4,5}$ *original*	$\{S_i\}_{i=1,2,3,4,5}$ *generated*
A_1	0.8184	0.8614	0.7685	0.7155	0.7917	0.7950
A_2	0.7928	0.7499	0.7216	0.6586	0.7576	0.7007
$\{A_1, A_2\}$	0.8034	0.7851	0.7271	0.6696	0.7664	0.7247

AImage Lab

Conclusions

- The detail accuracy of the proposed model is quite good, compared with the tested competitors;
- δ-*metrics* commonly used ($\delta < 1.25, \delta < 1.25^2, \delta < 1.25^3$), are effective to check the overall quality of depth maps generated from landscapes or wide-angle scenes, but **the threshold value is too high to take fine details into account**;
- We introduce **a new set of *δ-metrics*** ($\delta < 1.25^{\frac{1}{2}}, \delta < 1.25^{\frac{1}{3}}, \delta < 1.25^{\frac{1}{4}}$) with harder thresholds

Detail accuracy is still an open problem
in the **depth estimation** task with **Conditional GANs**

We note that our approach is able to produce overall accurate views of the generated facial depth maps:

- preserves the **shape of the face**
- maintain **garment details**