Video anonymization

Prof. Dr. Laura Leal-Taixé

Technical University of Munich

“All human beings have three lives: public, private, and secret.”

Gabriel García Márquez
Motivation

How I see my work

• Challenging
• Plenty of applications: autonomous driving, robot navigation

Data from www.motchallenge.net

How others see my work

Big brother
Motivation

How I see my work

- Challenging
- Plenty of applications: autonomous driving, robot navigation

How others see my work

I do not care if this is Mark or John, I only use a label “person”

Data from www.motchallenge.net
Motivation

Just remove a face using blur/square/mosaic

Motivation

Detection and tracking performance is heavily affected.

Images:
Goals for anonymization

Properties:
- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control
- Temporal Consistency

Reference:
Face swap

Properties:

- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control
- Temporal Consistency

Reference:
Properties:

- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control
- Temporal Consistency
Anonymization: previous work

Properties:

- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control (one-to-many)
- Temporal Consistency
Who is he?

More anonymized  ←  Less anonymized

Gafni et al. “Live face de-identification in video”. ICCV 2019

M. Maximov et al. „CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks“. CVPR 2020
Anonymization: previous work

Properties:

- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control (one-to-many)
- Temporal Consistency
CIAGAN

- Anonymous
- Realistic
- New Identity
- Control
- Temporal Consistency

Reference:
CIAGAN

CNN

Control over identity

Person/Full body

- Also works on full bodies!

Reference:
Methodology
Overview of CIAGAN

**Control over identity**

**MLP**

**Landmark detection**

**CNN**

**Output / Fake**

**Input**

**Shape + Background**

**Control over identity**
Inputs

- Partial Landmarks
  - We do not want appearance of the input face to "leak" to the new face
  - Mouth for expressions
  - Nose & Frame for orientation
  - "Free" temporal consistency

- Background Image
  - From Landmarks
  - For better blending of the face with the head and hair
Without further losses, the network overfits and simply does reconstruction.
Losses 2: ID Loss

Input

Landmark detection

Shape + Background

Training set

CNN

Output / Fake

Discriminator

Real / Fake

Identity Discr.

ID Embeddings

Control over identity

0 1 0 ... 0
Identity Guidance

- **Input:**
  - One-hot vector encoding of a random ID of the training set
  - We pass it through an MLP and obtain a representation which is then concatenated at the bottleneck of the CNN

- **Decoder:**
  - Effectively uses the encoded information of the initial ID and mixes it with one of the random training IDs

---

In how many ways can we anonymize an image?
• **Identity Discriminator**
  - Pre-train for re-ID on real images with Proxy-NCA loss
  - Contrastive loss during GAN training: brings the embedding of the new ID closer to the real training ID embedding
The identity discriminator is not used as adversarial, is it a guidance for the generator.
And for multi-object tracking?

- At each frame of a video:
  - We apply the *same transformation* to all pedestrians, so that we can perform tracking across frames.

- For a different camera
  - We apply the *a different transformation* to avoid long-term tracking and potential misuse of the data.
Results
Qualitative results

Control

identity

Source
## Detection & Identification

Detection and identification on the CelebA dataset

<table>
<thead>
<tr>
<th>Models</th>
<th>Detection (↑)</th>
<th>Identification (↓)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dlib</td>
<td>SSH</td>
</tr>
<tr>
<td>Original</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Pixelization 16 by 16</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Pixelization 8 by 8</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Blur 9 by 9</td>
<td>90.6</td>
<td>38.6</td>
</tr>
<tr>
<td>Blur 17 by 17</td>
<td>68.4</td>
<td>0.3</td>
</tr>
<tr>
<td>Ours</td>
<td>99.9</td>
<td>98.7</td>
</tr>
</tbody>
</table>

Blurring

Pixelization
## Ablation studies

<table>
<thead>
<tr>
<th>Models</th>
<th>Detection (↑)</th>
<th>Recall@1 (↓)</th>
<th>FID (↓)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siamese</td>
<td><strong>99.9</strong></td>
<td>1.3</td>
<td>2.1</td>
</tr>
</tbody>
</table>

- **Face detection**
- **Identification**
- **Visual quality**
Ablation studies

- Classification of the Identity instead of Siamese training:
  - Identity recall goes down, mostly because the generated faces start to have artifacts → low detection rate and poor visual quality

<table>
<thead>
<tr>
<th>Models</th>
<th>Detection (↑)</th>
<th>Recall @1 (↓)</th>
<th>FID (↓)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siamese</td>
<td>99.9</td>
<td>1.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Classification</td>
<td>64.6</td>
<td>0.4</td>
<td>63.2</td>
</tr>
</tbody>
</table>
Ablation studies

- Input are full face images instead of landmarks.
  - Visual quality of the generated faces and detectability both decrease

<table>
<thead>
<tr>
<th>Models</th>
<th>Detection (↑)</th>
<th>Recall@1 (↓)</th>
<th>FID (↓)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siamese</td>
<td>99.9</td>
<td>1.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Classification</td>
<td>64.6</td>
<td>0.4</td>
<td>63.2</td>
</tr>
<tr>
<td>Faces</td>
<td>98.3</td>
<td>1.1</td>
<td>6.5</td>
</tr>
</tbody>
</table>
Comparison with SOA

Two methods for face identification

<table>
<thead>
<tr>
<th>De-ID method</th>
<th>VGGFace2 (↓)</th>
<th>CASIA (↓)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.986 ±0.010</td>
<td>0.965 ±0.016</td>
</tr>
<tr>
<td>Gafni et al.</td>
<td>0.038 ±0.015</td>
<td>0.035 ±0.011</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>0.029 ± 0.012</strong></td>
<td><strong>0.026 ± 0.015</strong></td>
</tr>
</tbody>
</table>

- We are able to mask identities better
  - While also providing more diversity in the output and more control
Comparison with SOA

Anonymization variations

- We are able to mask identities better
  - While also providing more diversity in the output and more control
Glasses & Hair & Makeup

Source Anonymizations

Source Anonymization
Results

Source | Anonymizations

[Images of anonymized faces]
Different Domain
Limitations

Source | Part to replace | Landmark | Background | Result
---|---|---|---|---

- **Extreme Poses**
  - Eyes
Future Work

- Occlusions
- Different Domains
- Study the effect on multiple object tracking
- Do not depend on the output of the landmarks
- More realistic and high-definition images
- Work on explicit temporal consistency
The Team

Maxim Maximov  Ismail Elezi  Laura Leal-Taixé
“All human beings have three lives: public, private, and secret.”

Gabriel García Márquez