UNIVERSITY OF BUCHAREST

VIRTUTE ET SAPIENTIA

# Teacher-Student Training and Triplet Loss for Facial Expression Recognition under Occlusion

Mariana-Iuliana Georgescu[1, 2], Radu Tudor Ionescu[1]

[1]University of Bucharest, Bucharest, Romania,
[2]Novustech Services, Bucharest, Romania


# Motivation

- We focus on the application of facial expression recognition to human-computer interaction in a virtual reality environment

- Our aim is to design a system:
  - to recognize the facial expressions of a user wearing a VR headset
  - to automatically control and adjust the VR environment according to the user's emotions

ICPR 2020

# Related work

- synthetically occluded images
  [Li et al. ICPR 2018]
- analyzes expressions from the eye region
  [Hickson et al. WACV 2019]
- analyzes the mouth region captured with a standard camera
  [Georgescu et al. ICONIP 2019]

# Our method

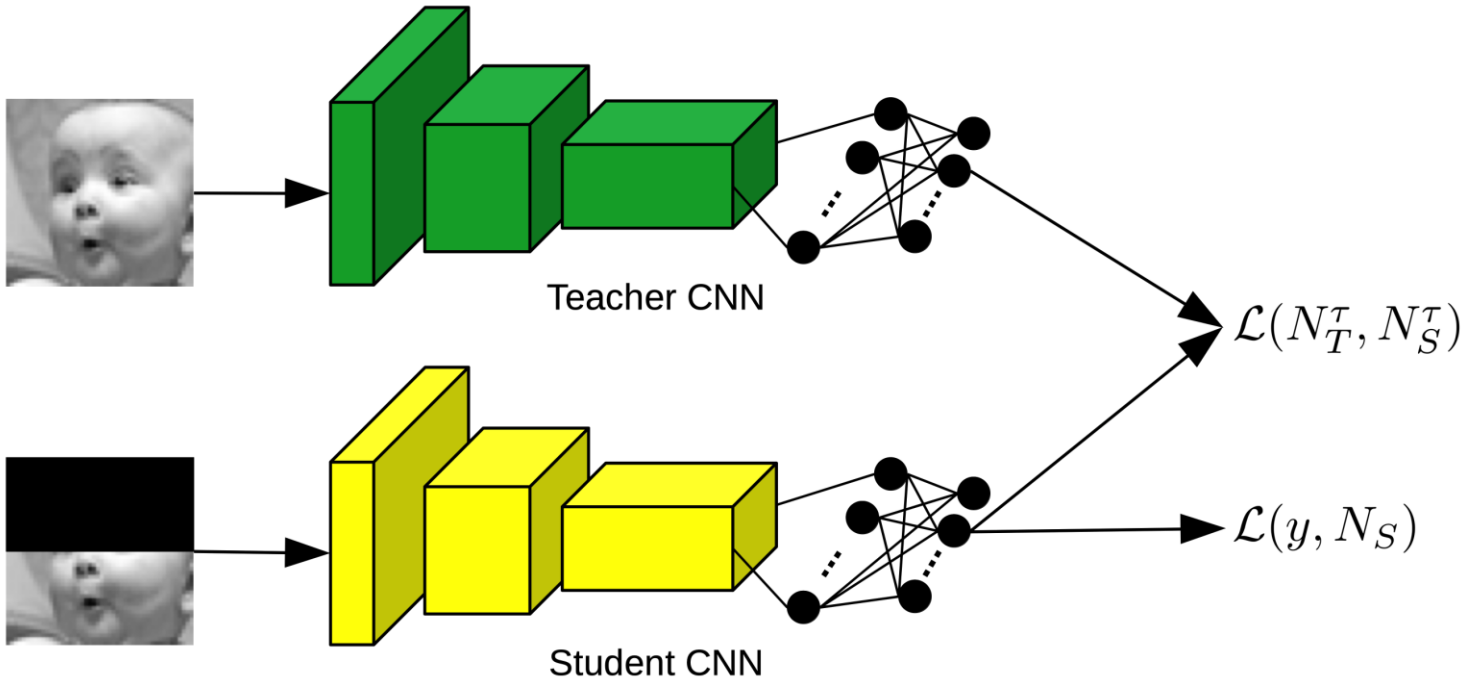
- Our approach is based on the following steps:

1. Fine-tuning a teacher CNN on full faces
2. Fine-tuning a student CNN on lower-half faces
3. Knowledge distillation

# Our method

- We consider two knowledge distillation methods to obtain more accurate CNNs based on:
  - teacher-student training
  - triplet loss (novel)

# Teacher-student



# Teacher-student

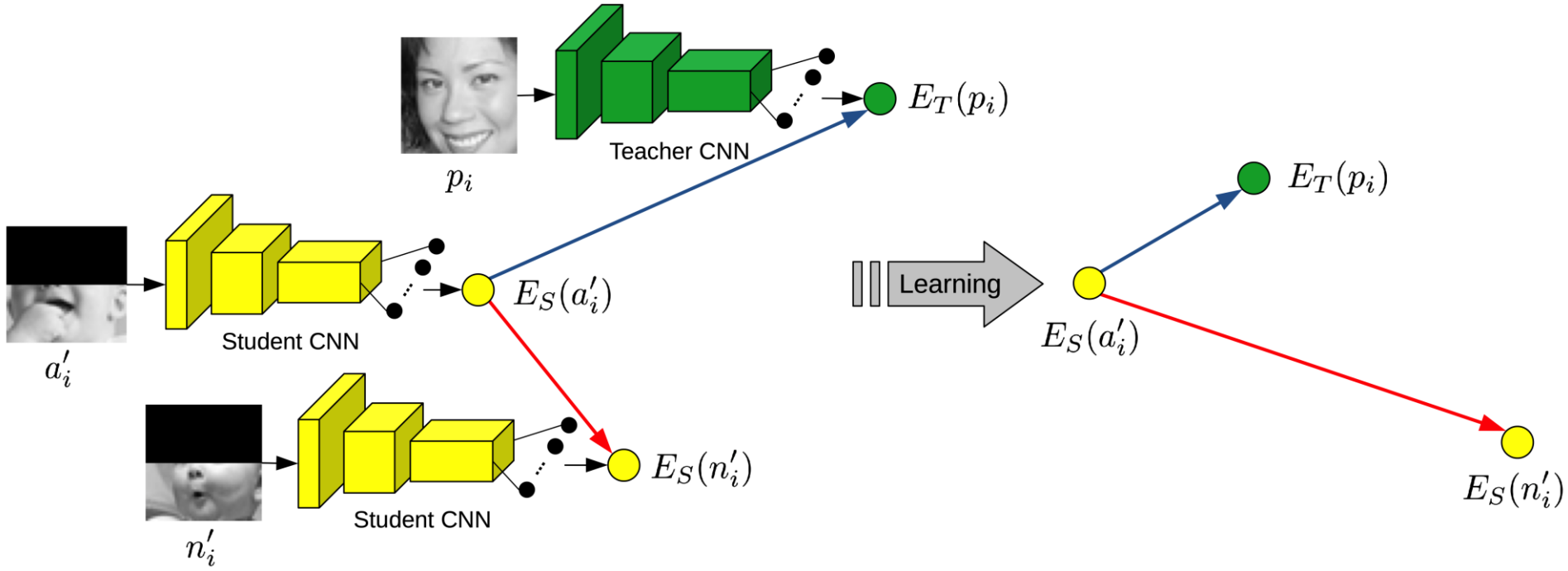
Knowledge distillation loss function:

$$\mathcal{L}_{KD'}(\theta_S) = (1-\lambda)\mathcal{L}(y, N_S) + \lambda\mathcal{L}(N_T^\tau, N_S^\tau)$$

where:

$$N_T^\tau = softmax\left(\frac{A_T}{\tau}\right), N_S^\tau = softmax\left(\frac{A_S}{\tau}\right)$$

# Triplet loss



# Triplet loss

Triplet loss function for knowledge distillation:

$$\mathcal{L}_{KD''}(\theta_S) = \mathcal{L}(y, N_S) + \lambda \mathcal{L}_{triplet}(\theta_S)$$

where:

$$\mathcal{L}_{triplet}(\theta_S) = \sum_{i=1}^{m} \Big[ \Delta(E_S(a'_i), E_T(p_i)) - \Delta(E_S(a'_i), E_S(n'_i)) + \alpha \Big]_+$$

$E_T(x)$ and $E_S(x)$ are the embeddings produced by the teacher T and the student S, $\Delta$ is the Euclidian distance

# Experiments

- Data sets:
  - FER+
  - AffectNet

| Data set | #training | #validation | #test |
|---|---|---|---|
| FER+ | 25045 | 3191 | 3137 |
| AffectNet | 287651 | 2000 | 4000 |

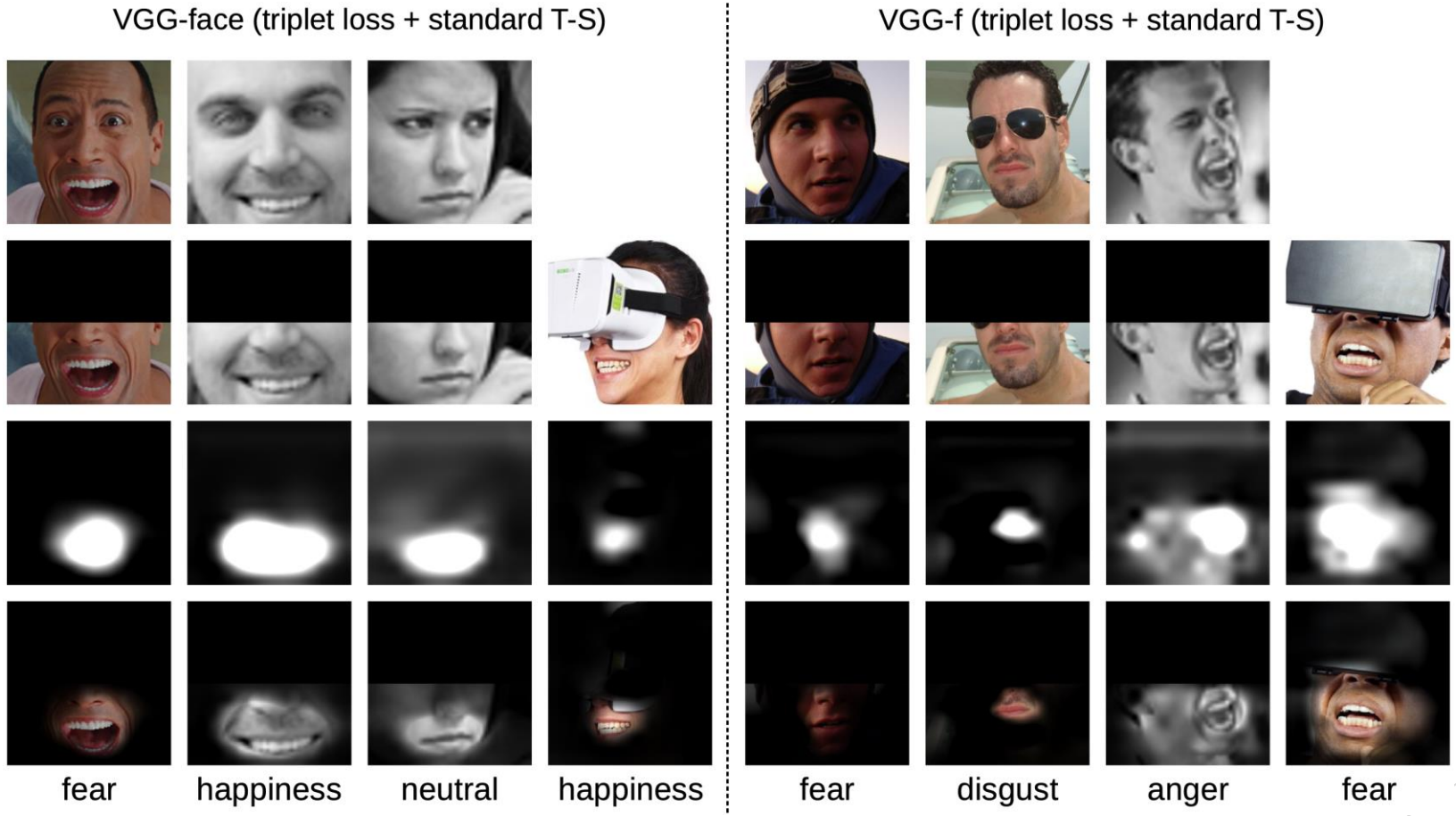
- CNN architectures
  - VGG-face (pre-trained for face recognition)
  - VGG-f (pre-trained on ImageNet)

# Results

| Model | Test faces | AffectNet | FER+ |
|---|---|---|---|
| Bag-of-visual-words [9] | ○ | 48.30% | 80.65% |
| VGG-13 [6] | ○ | - | 84.99% |
| AlexNet [7] | ○ | 58.00% | - |
| MT-VGG [10] | ○ | 54.00% | - |
| VGG-12 [8] | ○ | 58.50% | - |
| Teacher VGG-f | ○ | 57.37% | 85.05% |
| Teacher VGG-face | ○ | 59.03% | 84.79% |
| Teacher VGG-f | ◓ | 41.58% | 70.00% |
| Teacher VGG-face | ◓ | 37.70% | 68.89% |
| Teacher VGG-f | ◒ | 26.85% | 40.07% |
| Teacher VGG-face | ◒ | 31.23% | 48.29% |
| VGG-f [1] | ◓ | 47.58% | 78.23% |
| VGG-face [1] | ◓ | 49.23% | 81.28% |
| VGG-f [2] | ◒ | 42.45% | 66.18% |
| VGG-face [2] | ◒ | 43.18% | 70.19% |
| VGG-f (standard T-S) | ◓ | 48.75%† | 80.17%† |
| VGG-face (standard T-S) | ◓ | 49.75% | 82.37% |
| VGG-f (triplet loss T-S) | ◓ | 48.13% | 80.05%† |
| VGG-face (triplet loss T-S) | ◓ | 49.71% | 82.57% |
| VGG-f (triplet loss + standard T-S) | ◓ | 48.70%† | 81.09%† |
| VGG-face (triplet loss + standard T-S) | ◓ | 50.09%† | 82.75%† |

Accuracy rates of various models on AffectNet and FER+, for fully-visible faces (denoted by ○), lower-half-visible faces (denoted by ◓) and upper-half-visible faces (denoted by ◒)

# Grad-cam activation maps



# Conclusions

- We proposed to train neural networks for facial expression recognition under strong occlusion, by applying two knowledge distillation strategies.

- On FER+, our VGG-face based on concatenated distilled embeddings attains an accuracy rate of 82.75% on lower-half-visible faces, which is only 2.24% below the accuracy rate of the state-of-the-art VGG-13 on fully-visible faces.

ICPR 2020

UNIVERSITY OF BUCHAREST

VIRTUTE ET SAPIENTIA

# Teacher-Student Training and Triplet Loss for Facial Expression Recognition under Occlusion

Mariana-Iuliana Georgescu[1, 2], Radu Tudor Ionescu[1]

[1]University of Bucharest, Bucharest, Romania,
[2]Novustech Services, Bucharest, Romania