Deep Facial Expression Recognition Exploiting Categorical and Continuous Emotional Dependencies

Panagiotis Antoniadis

Division of Signals, Control and Robotics
School of Electrical and Computer Engineering, NTUA

Athens, November 2021
# Table of Contents

1. Introduction

2. Baseline Architecture

3. Metric Learning

4. Multi-task Learning

5. Emotion-GCN

6. Conclusion and Future Work
Introduction
Introduction

Panagiotis Antoniadis (ECE NTUA)  Facial Expression Recognition  Athens, November 2021 4 / 36
Introduction

Design and train the FER system

Choose an emotion representation

Panagiotis Antoniadis (ECE NTUA)
Facial Expression Recognition
Athens, November 2021
1. Design and train the FER system
Introduction

1. Design and train the FER system
2. Choose an emotion representation
Categorical Model

Primary and universal emotions proposed by Paul Ekman

- Happy
- Sad
- Surprise
- Fear
- Anger
- Disgust
- Neutral

- Simple and intuitive
- Restricts emotion in discrete categories
Introduction

Dimensional Model (or VA)

Continuous model along a set of 2 dimensions (valence/arousal)

✓ Describes complex and subtle emotions

× Annotation is challenging
Introduction

FER → Lab

→ In the wild

SOLVED

NOT YET
Introduction

Why is FER in-the-wild challenging?

FER

Lab   SOLVED

In the wild    NOT YET

Why is FER in-the-wild challenging?
Introduction

Why is FER in-the-wild challenging?

- Identity bias (age, gender, hair)
- Head pose variations
- Illumination variations
- Occlusions
- Subjectivity / Inherent variation
Introduction

Why is FER in-the-wild challenging?

- Identity bias (age, gender, hair)
- Head pose variations
- Illumination variations
- Occlusions
- Subjectivity / Inherent variation
Introduction

Why is FER in-the-wild challenging?

- Identity bias (age, gender, hair)
- Head pose variations
- Illumination variations
- Occlusions
- Subjectivity / Inherent variation

FER \[\rightarrow\] Lab SOLVED

FER \[\rightarrow\] In the wild NOT YET
Introduction

Why is FER in-the-wild challenging?

- Identity bias (age, gender, hair)
- Head pose variations
- Illumination variations
- Occlusions
- Subjectivity / Inherent variation

FER \[\rightarrow\text{Lab} \quad \text{SOLVED}\]

FER \[\rightarrow\text{In the wild} \quad \text{NOT YET}\]
Introduction

Why is FER in-the-wild challenging?

- Identity bias (age, gender, hair)
- Head pose variations
- Illumination variations
- Occlusions
- Subjectivity / Inherent variation

FER

Lab  SOLVED

In the wild  NOT YET
Introduction

Our contribution

✓ Train metric learning models to reduce the variations of the task.
Introduction

Our contribution

✓ Train metric learning models to reduce the variations of the task.

✓ Train multi-task learning models to explore the relation between the emotion representations.
Introduction

Our contribution

✓ Train metric learning models to reduce the variations of the task.

✓ Train multi-task learning models to explore the relation between the emotion representations.

✓ Propose Emotion-GCN that uses a Graph Convolutional Network to capture the emotional dependencies.
Introduction

Our contribution

✓ Train metric learning models to reduce the variations of the task.

✓ Train multi-task learning models to explore the relation between the emotion representations.

✓ Propose Emotion-GCN that uses a Graph Convolutional Network to capture the emotional dependencies.

Our proposed model achieves state-of-the-art results on AffectNet dataset, the largest in-the-wild database of facial expressions.
Table of Contents

1 Introduction

2 Baseline Architecture

3 Metric Learning

4 Multi-task Learning

5 Emotion-GCN

6 Conclusion and Future Work
Baseline Architecture

Preprocessing → Deep Architecture → Loss Function → Emotion
Baseline Architecture

- Preprocessing
- Deep Architecture
- Loss Function
- Emotion
Baseline Architecture

1. Face Detection
Baseline Architecture

1. Face Detection
2. Face Alignment
Baseline Architecture

We use a Densely Connected Convolutional Network (DenseNet).

- Tackles the vanishing gradients problem
- Parameter-efficient
Baseline Architecture

Categorical

$$- \sum_{i=1}^{7} \frac{f_i}{f_{\text{min}}} \ y_i \log(\hat{y}_i)$$

$$\hat{y}_i$$: probability of emotion $$i$$
$$y_i = 1$$ if emotion $$i$$ is the label
$$f_i$$: number of samples in emotion $$i$$
$$f_{\text{min}}$$: number of samples in the most under-represented class

Dimensional

$$1 - \frac{\rho_v + \rho_a}{2}$$

$$\rho_v, \rho_a$$: CCC of valence and arousal

$$\rho_c = \frac{2s_{xy}}{s_x^2 + s_y^2 + (\bar{x} - \bar{y})^2}$$

$$s_x$$: variance of labels
$$s_y$$: variance of predictions
$$s_{xy}$$: covariance
$$\bar{x}, \bar{y}$$: mean values
Metric Learning

Appearance
Metric Learning

Appearance

Intra-class

Anger

Anger
Metric Learning

Appearance

Intra-class

Inter-class

Anger

Anger

Anger

Disgust
Metric Learning

Appearance

Intra-class

Anger

Anger

Inter-class

Anger

Disgust

Feature Extraction

$I \rightarrow f_{\text{cnn}} \rightarrow f_{\text{gmp}} \rightarrow x \rightarrow L_{\text{softmax}}$

$L_{\text{extra}}$

Neutral

Happy

Sad

Surprise

Fear

Anger

Disgust

Panagiotis Antoniadis (ECE NTUA)

Facial Expression Recognition

Athens, November 2021
Metric Learning

**Appearance**

**Intra-class**

**Inter-class**

Goal → Apply $L_{extra}$ to reduce the impact of variations
**Center Loss:** Reduce intra-class variation of learned features
Metric Learning

**Center Loss:** Reduce intra-class variation of learned features

\[ L_c = \frac{1}{2} \sum_{i=1}^{m} \| x_i - c_{y_i} \|_2^2 \]

- \( x_i \): feature vector of sample \( i \).
- \( y_i \): class of sample \( i \).
- \( c_{y_i} \): center of class \( y_i \).
**Center Loss:** Reduce intra-class variation of learned features

\[ L_c = \frac{1}{2} \sum_{i=1}^{m} ||x_i - c_{y_i}||^2 \]

- \( x_i \): feature vector of sample \( i \).
- \( y_i \): class of sample \( i \).
- \( c_{y_i} \): center of class \( y_i \).

Softmax only (left) and softmax with center loss (right)
Extensions of Center Loss

- **Island Loss**: Increase inter-class differences

\[
L_{island} = L_c + \lambda_1 \sum_{c_j \in N} \sum_{c_k \in N, c_k \neq c_j} \left( \frac{c_k \cdot c_j}{||c_k||_2 ||c_j||_2} + 1 \right)
\]
 Extensions of Center Loss

- **Island Loss**: Increase inter-class differences

  \[ L_{island} = L_c + \lambda_1 \sum_{c_j \in N} \sum_{c_k \in N} \sum_{c_k \neq c_j} \left( \frac{c_k \cdot c_j}{\|c_k\|_2 \|c_j\|_2} + 1 \right) \]

- **Local Subclass Loss**: Many centers for each class

  \[ L_{subclass} = \frac{1}{2} \sum_{i=1}^{m} \|x_i - c_{y_i}^{min}\|_2^2 \]
Metric Learning

Extensions of Center Loss

- **Island Loss:** Increase inter-class differences

\[ L_{\text{island}} = L_c + \lambda_1 \sum_{c_j \in N} \sum_{c_k \in N, c_k \neq c_j} \left( \frac{c_k \cdot c_j}{\|c_k\|_2 \|c_j\|_2} + 1 \right) \]

- **Local Subclass Loss:** Many centers for each class

\[ L_{\text{subclass}} = \frac{1}{2} \sum_{i=1}^{m} \|x_i - c_{y_i}^{\text{min}}\|_2^2 \]

- **VA-based Center Loss:** Use VA values

\[ L_{c}^{\text{va}} = \frac{1}{2} \sum_{i=1}^{m} w_i \|x_i - c_{y_i}\|_2^2 \]
**Metric Learning**

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>64.37</td>
</tr>
<tr>
<td>Center loss ($\lambda = 0.1$)</td>
<td>64.37</td>
</tr>
<tr>
<td>Island loss ($\lambda = 0.1$)</td>
<td>65.11</td>
</tr>
<tr>
<td>Local subclass loss with margin ($\lambda = 0.1$, $k=10$, $m=0.5$)</td>
<td><strong>65.29</strong></td>
</tr>
<tr>
<td>VA-based Center local loss</td>
<td>64.91</td>
</tr>
</tbody>
</table>

Panagiotis Antoniadis (ECE NTUA)  
Facial Expression Recognition  
Athens, November 2021
**Metric Learning**

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>64.37</td>
</tr>
<tr>
<td>Center loss ($\lambda = 0.1$)</td>
<td>64.37</td>
</tr>
<tr>
<td>Island loss ($\lambda = 0.1$)</td>
<td>65.11</td>
</tr>
<tr>
<td>Local subclass loss with margin ($\lambda = 0.1$, $k=10$, $m=0.5$)</td>
<td><strong>65.29</strong></td>
</tr>
<tr>
<td>VA-based Center local loss</td>
<td>64.91</td>
</tr>
</tbody>
</table>

- Metric learning models outperform the baseline.
## Metric Learning

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>64.37</td>
</tr>
<tr>
<td>Center loss ($\lambda = 0.1$)</td>
<td>64.37</td>
</tr>
<tr>
<td>Island loss ($\lambda = 0.1$)</td>
<td>65.11</td>
</tr>
<tr>
<td>Local subclass loss with margin ($\lambda = 0.1$, $k=10$, $m=0.5$)</td>
<td><strong>65.29</strong></td>
</tr>
<tr>
<td>VA-based Center local loss</td>
<td>64.91</td>
</tr>
</tbody>
</table>

- Metric learning models outperform the baseline.
- Learned vectors are more discriminative in the feature space.
## Metric Learning

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>64.37</td>
</tr>
<tr>
<td>Center loss ($\lambda = 0.1$)</td>
<td>64.37</td>
</tr>
<tr>
<td>Island loss ($\lambda = 0.1$)</td>
<td>65.11</td>
</tr>
<tr>
<td>Local subclass loss with margin ($\lambda = 0.1$, $k=10$, $m=0.5$)</td>
<td><strong>65.29</strong></td>
</tr>
<tr>
<td>VA-based Center local loss</td>
<td>64.91</td>
</tr>
</tbody>
</table>

- Metric learning models outperform the baseline.
- Learned vectors are more discriminative in the feature space.

**But we ignore the dimensional emotion representation!**
Table of Contents

1 Introduction
2 Baseline Architecture
3 Metric Learning
4 Multi-task Learning
5 Emotion-GCN
6 Conclusion and Future Work
Multi-task Learning

- Use the knowledge for one task to aid the learning of another task.
- Learn more robust and universal representations.
Multi-task Learning

- Use the knowledge for one task to aid the learning of another task.
- Learn more robust and universal representations.

Distribution of the basic expressions in the VA space that illustrates the emotional dependencies between the categorical and the dimensional model.
Multi-task Learning

- Multi-task network on the categorical and the dimensional model.
Multi-task Learning

Convert the regression task to a classification task → Divide the VA space in regions.

✓ Angular division
✓ 4, 8 or 12 regions
Multi-task Learning

- Convert the regression task to a classification task \( \rightarrow \) Divide the VA space in regions.

![Diagram of Feature Extraction and Classification](image)
## Multi-task Learning

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>64.37</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>64.8</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td>65.63</td>
</tr>
<tr>
<td>Multi-task + Quarter</td>
<td>65.14</td>
</tr>
<tr>
<td><strong>Multi-task + Eight</strong></td>
<td><strong>65.86</strong></td>
</tr>
<tr>
<td>Multi-task + Twelve</td>
<td>65.11</td>
</tr>
</tbody>
</table>

Multi-task models outperform the single-task model. Our main task benefits from the integration of the VA values. The CCC loss performs better than MSE. Dividing the VA space is more effective than regression.
## Multi-task Learning

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>64.37</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>64.8</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td>65.63</td>
</tr>
<tr>
<td>Multi-task + Quarter</td>
<td>65.14</td>
</tr>
<tr>
<td><strong>Multi-task + Eight</strong></td>
<td><strong>65.86</strong></td>
</tr>
<tr>
<td>Multi-task + Twelve</td>
<td>65.11</td>
</tr>
</tbody>
</table>

- Multi-task models outperform the single-task model.
Multi-task Learning

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>64.37</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>64.8</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td>65.63</td>
</tr>
<tr>
<td>Multi-task + Quarter</td>
<td>65.14</td>
</tr>
<tr>
<td><strong>Multi-task + Eight</strong></td>
<td><strong>65.86</strong></td>
</tr>
<tr>
<td>Multi-task + Twelve</td>
<td>65.11</td>
</tr>
</tbody>
</table>

- Multi-task models outperform the single-task model.
- Our main task benefits from the integration of the VA values.
## Multi-task Learning

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>64.37</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>64.8</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td>65.63</td>
</tr>
<tr>
<td>Multi-task + Quarter</td>
<td>65.14</td>
</tr>
<tr>
<td><strong>Multi-task + Eight</strong></td>
<td><strong>65.86</strong></td>
</tr>
<tr>
<td>Multi-task + Twelve</td>
<td>65.11</td>
</tr>
</tbody>
</table>

- Multi-task models outperform the single-task model.
- Our main task benefits from the integration of the VA values.
- The CCC loss performs better than MSE.
## Multi-task Learning

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>64.37</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>64.8</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td>65.63</td>
</tr>
<tr>
<td>Multi-task + Quarter</td>
<td>65.14</td>
</tr>
<tr>
<td><strong>Multi-task + Eight</strong></td>
<td><strong>65.86</strong></td>
</tr>
<tr>
<td>Multi-task + Twelve</td>
<td>65.11</td>
</tr>
</tbody>
</table>

- Multi-task models outperform the single-task model.
- Our main task benefits from the integration of the VA values.
- The CCC loss performs better than MSE.
- Dividing the VA space is more effective than regression.
<table>
<thead>
<tr>
<th>#</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Baseline Architecture</td>
</tr>
<tr>
<td>3</td>
<td>Metric Learning</td>
</tr>
<tr>
<td>4</td>
<td>Multi-task Learning</td>
</tr>
<tr>
<td>5</td>
<td>Emotion-GCN</td>
</tr>
<tr>
<td>6</td>
<td>Conclusion and Future Work</td>
</tr>
</tbody>
</table>
Emotion-GCN

The strong dependence between the categorical and the dimensional model is not fully exploited when they only share a feature representation.
The strong dependence between the categorical and the dimensional model is not fully exploited when they only share a feature representation.
The strong dependence between the categorical and the dimensional model is not fully exploited when they only share a feature representation.
Emotion-GCN

Multi-label image recognition

- Objects co-occur in the world
- Model label dependencies
- Learn dependent classifiers
Emotion-GCN

ML-GCN model for multi-label image recognition.

Graph Convolutional Network (GCN)

Generated classifiers

ML-GCN model for multi-label image recognition.
Emotion-GCN

Multi-task FER

- Strong dependencies between the emotion representations
- Model emotional dependencies
- Learn dependent expression classifiers and VA regressors
Emotion-GCN

Multi-task FER

1. Define the nodes of the graph

![Emotion Graph]

- Fear
- Sad
- Neutral
- Valence
- Arousal
- Surprise
- Happy
- Disgust
- Anger
**Emotion-GCN**

**Multi-task FER**

2 Design the adjacency matrix

\[
A_{ij} = \begin{cases} 
1, & \text{if } i = j \\
|c_{ij}|, & \text{if } i \in \text{Cat} \land j \in \text{Dim} \\
|c_{ij}|, & \text{if } j \in \text{Cat} \land i \in \text{Dim} \\
0, & \text{else}
\end{cases}
\]

\[
A'_{ij} = \begin{cases} 
1, & \text{if } A_{ij} \geq \tau \\
0, & \text{if } A_{ij} < \tau
\end{cases}
\]

\[
A''_{ij} = \begin{cases} 
(p/ \sum_{i\neq j}^{9} A'_{ij}) \times A'_{ij}, & \text{if } i \neq j \\
1 - p, & \text{if } i = j
\end{cases}
\]
## Multi-task FER

### Design the adjacency matrix

<table>
<thead>
<tr>
<th></th>
<th>Neutral</th>
<th>Happy</th>
<th>Sad</th>
<th>Surprise</th>
<th>Fear</th>
<th>Disgust</th>
<th>Anger</th>
<th>Valence</th>
<th>Arousal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
<td>0.35</td>
<td>0.30</td>
</tr>
<tr>
<td>Happy</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.70</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Sad</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
<td>0.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.70</td>
<td>0.00</td>
</tr>
<tr>
<td>Fear</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
<td>0.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.35</td>
<td>0.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Anger</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.35</td>
<td>0.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Valence</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.00</td>
<td>0.12</td>
<td>0.12</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Arousal</td>
<td>0.12</td>
<td>0.00</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Multi-task FER

Emotion-GCN model for facial expression recognition.
## Emotion-GCN

<table>
<thead>
<tr>
<th>Method</th>
<th>AffectNet</th>
<th>Aff-Wild2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>64.37</td>
<td>45.06</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>64.8</td>
<td>43.1</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td>65.69</td>
<td>43.33</td>
</tr>
<tr>
<td>Emotion-GCN</td>
<td><strong>66.46</strong></td>
<td><strong>48.92</strong></td>
</tr>
</tbody>
</table>

Categorical model
## Emotion-GCN

<table>
<thead>
<tr>
<th>Method</th>
<th>AffectNet</th>
<th>Aff-Wild2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>64.37</td>
<td>45.06</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>64.8</td>
<td>43.1</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td>65.69</td>
<td>43.33</td>
</tr>
<tr>
<td>Emotion-GCN</td>
<td><strong>66.46</strong></td>
<td><strong>48.92</strong></td>
</tr>
</tbody>
</table>

Categorical model

<table>
<thead>
<tr>
<th>Method</th>
<th>AffectNetCCC-V</th>
<th>AffectNetCCC-A</th>
<th>Aff-Wild2CCC-V</th>
<th>Aff-Wild2CCC-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-task</td>
<td>0.761</td>
<td>0.628</td>
<td>0.416</td>
<td>0.501</td>
</tr>
<tr>
<td>Multi-task + MSE</td>
<td>0.752</td>
<td>0.572</td>
<td>0.435</td>
<td>0.378</td>
</tr>
<tr>
<td>Multi-task + CCC</td>
<td><strong>0.768</strong></td>
<td><strong>0.651</strong></td>
<td>0.408</td>
<td>0.481</td>
</tr>
<tr>
<td>Emotion-GCN</td>
<td>0.767</td>
<td>0.649</td>
<td><strong>0.457</strong></td>
<td><strong>0.514</strong></td>
</tr>
</tbody>
</table>

Dimensional model
Emotion-GCN

Panagiotis Antoniadis (ECE NTUA)  Facial Expression Recognition  Athens, November 2021  30 / 36
## Emotion-GCN

Predictions of Emotion-GCN on samples of AffectNet.
Emotion-GCN

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA2LT</td>
<td>57.31</td>
</tr>
<tr>
<td>gACNN</td>
<td>58.78</td>
</tr>
<tr>
<td>Facial Motion Prior Network</td>
<td>61.52</td>
</tr>
<tr>
<td>CAKE</td>
<td>61.7</td>
</tr>
<tr>
<td>OADN</td>
<td>61.89</td>
</tr>
<tr>
<td>CNNs and BOVW + global SVM</td>
<td>63.31</td>
</tr>
<tr>
<td>Siamese</td>
<td>64</td>
</tr>
<tr>
<td><strong>Emotion-GCN (ours)</strong></td>
<td><strong>66.46</strong></td>
</tr>
</tbody>
</table>

Comparison with SOTA on AffectNet
Comparison with SOTA on AffectNet

- Significant improvements in both the categorical and the VA model.
## Emotion-GCN

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA2LT</td>
<td>57.31</td>
</tr>
<tr>
<td>gACNN</td>
<td>58.78</td>
</tr>
<tr>
<td>Facial Motion Prior Network</td>
<td>61.52</td>
</tr>
<tr>
<td>CAKE</td>
<td>61.7</td>
</tr>
<tr>
<td>OADN</td>
<td>61.89</td>
</tr>
<tr>
<td>CNNs and BOVW + global SVM</td>
<td>63.31</td>
</tr>
<tr>
<td>Siamese</td>
<td>64</td>
</tr>
<tr>
<td><strong>Emotion-GCN (ours)</strong></td>
<td><strong>66.46</strong></td>
</tr>
</tbody>
</table>

Comparison with SOTA on AffectNet

- Significant improvements in both the categorical and the VA model.
- The benefits of our proposed model generalize in more datasets.
### Emotion-GCN

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA2LT</td>
<td>57.31</td>
</tr>
<tr>
<td>gACNN</td>
<td>58.78</td>
</tr>
<tr>
<td>Facial Motion Prior Network</td>
<td>61.52</td>
</tr>
<tr>
<td>CAKE</td>
<td>61.7</td>
</tr>
<tr>
<td>OADN</td>
<td>61.89</td>
</tr>
<tr>
<td>CNNs and BOVW + global SVM</td>
<td>63.31</td>
</tr>
<tr>
<td>Siamese</td>
<td>64</td>
</tr>
<tr>
<td><strong>Emotion-GCN (ours)</strong></td>
<td><strong>66.46</strong></td>
</tr>
</tbody>
</table>

Comparison with SOTA on AffectNet

- Significant improvements in both the categorical and the VA model.
- The benefits of our proposed model generalize in more datasets.
- The learned vectors effectively capture the emotional dependencies.
Emotion-GCN

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPA2LT</td>
<td>57.31</td>
</tr>
<tr>
<td>gACNN</td>
<td>58.78</td>
</tr>
<tr>
<td>Facial Motion Prior Network</td>
<td>61.52</td>
</tr>
<tr>
<td>CAKE</td>
<td>61.7</td>
</tr>
<tr>
<td>OADN</td>
<td>61.89</td>
</tr>
<tr>
<td>CNNs and BOVW + global SVM</td>
<td>63.31</td>
</tr>
<tr>
<td>Siamese</td>
<td>64</td>
</tr>
<tr>
<td>Emotion-GCN (ours)</td>
<td>66.46</td>
</tr>
</tbody>
</table>

Comparison with SOTA on AffectNet

- Significant improvements in both the categorical and the VA model.
- The benefits of our proposed model generalize in more datasets.
- The learned vectors effectively capture the emotional dependencies.
- **We achieve SOTA results on AffectNet.**
Table of Contents

1 Introduction
2 Baseline Architecture
3 Metric Learning
4 Multi-task Learning
5 Emotion-GCN
6 Conclusion and Future Work
Conclusion

Contributions

1. We proposed metric learning models to reduce variations.

2. We proposed multi-task learning models to exploit the categorical and continuous emotional dependencies.

3. We proposed Emotion-GCN that uses a Graph Convolution Network to capture the emotional dependencies.
Conclusion

Contributions

1. We proposed metric learning models to reduce variations.
2. We proposed multi-task learning models to exploit the categorical and continuous emotional dependencies.
3. We proposed Emotion-GCN that uses a Graph Convolution Network to capture the emotional dependencies.

Publications

1. ”Exploiting Emotional Dependencies with Graph Convolutional Networks for Facial Expression Recognition”, P Antoniadis, P Filntisis, P Maragos, *IEEE FG 2021*. [paper] [code]
2. Along with Ioannis Pikoulis we participated in ABAW2 competition: ”An audiovisual and contextual approach for categorical and continuous emotion recognition in-the-wild”, P Antoniadis, I Pikoulis, P Filntisis, P Maragos, *ICCVW 2021*. [paper] [code]
Future Work

What’s next?

1. Extend proposed techniques for videos.
2. Explore bias in annotations of emotion databases.
3. Use transformers in FER.
Future Work

What’s next?

1. Extend proposed techniques for videos.
2. Explore bias in annotations of emotion databases.
3. Use transformers in FER.
Thank You

Questions?