Trabalhos Relacionados

Todos os trabalhos citados trabalham com a base de dados FER2013 e **quase** todos aplicaram de alguma forma *ensemble* de redes neurais convolucionais para classificação estática, em apenas uma imagem.

Estas informações foram coletados de um survey deste ano, 23 de abril. No link: https://arxiv.org/pdf/1804.08348.pdf . E também diretamente dos artigos originais.

Principais diferenças para trabalhos relacionados

- FC pequenas

---

1. Facial Expression Recognition Using a Hybrid CNN-SIFT Aggregator

```
FER Using a Hybrid CNN - SIFT Aggregator
@article{al2016facial,
  title={Facial Expression Recognition Using a Hybrid CNN-SIFT Aggregator},
  author={Al-Shabi, Mundher and Cheah, Wooi Ping and Connie, Tee},
  journal={arXiv preprint arXiv:1608.02833},
  year={2016}
}
```

Trechos importantes do *paper*:

- Automatic facial expressions recognition (FER) has been an active research in the computer vision field. Facial expression and emotion recognition with hand-crafted feature extractors were reported in [2]–[5]. Many works have also applied convolution neural network in facial expression recognition. In [8] the authors analyzed the features learnt by neural network and showed that neural network could learn patterns from the face images that correspond to Facial Action Units (FAUs). He proposed to ignore the biases of the convolutional layers which gave him an accuracy of 98.3% on the CK+ dataset. The winner of FER-2013 challenge [22] used a CNN layer followed by a linear one-vs-all SVM. Instead of minimization of cross-entropy loss like vanilla CNN, he minimized a margin based loss with standard hinge loss. The method achieved 71.2% accuracy on a private test. Another study applied deeper neural network by constructing four inception layers after two ordinary convolution layers [18]. Models based on transfer features from pre-trained deep CNN have also been proposed [19, 25]. The importance of pre-processing on increasing the accuracy of the FER model was heavily studied by [14]. Different methods were used to increase the number of examples through rotation correction and intensity normalization.

- Besides, multiple CNN models were combined via learnable weights by minimizing the hinge loss [26]. The winner of EmotiW2015 [8] trained multiple CNNs as committee members and combined their decisions via construction of a hierarchical architecture of the committee with exponentially weighted decision fusion. The network architecture, input normalization, and random weight initialization were changed to obtain varied decisions for deep CNNs.

- In this paper, a hybrid Convolutional Neural Network with Dense Scale Invariant Feature Transform aggregator is proposed for facial expression recognition. We have shown how the Dense SIFT features and convolution neural network could complement each other in improving the accuracy result. The proposed method combines the strengths of hand-craft and deep learning approaches. The Dense SIFT technique is studied and compared with regular SIFT feature extractor in which it shows an advantage over regular SIFT. The performance gain is noticeable when all the models are combined with the aggregator in which it outperforms state-of-the-art methods. Our experiments demonstrate a clear advantage of aggregating Dense SIFT and CNN models by achieving outstanding results on both FER-2013 and CK+ datasets.

Contribuições

1. Combinação de SIFT e Dense SIFT com CNN
2. Multiplas CNN e SIFT

Acurácia de 73.4%

Pontos importantes

- Também usa data augmentation
  - Horizontal flip
  - Rotation with angle between(-30 to +30)
  - Skewing the center area
  - Zooming of four corners
- Standardization
- Arquitetura desenvolvida do zero
- Mesmo estilo de arquitetura(Maxpooling, kernel 3x3, strides 2x2, FC no final, dropout, softmax, função de ativação ReLu, otimizador adam)
- Para cálculo da expressão final é usado a média das resposta de CNN, CNN + SIFT, CNN + D-SIFT para cada expressão, a que tiver maior média é a expressão classificada

---

2. Facial Expression Recognition using Convolutional Neural Networks: State of the Art

```
@article{DBLP:journals/corr/PramerdorferK16,
  author    = {Christopher Pramerdorfer and
               Martin Kampel},
  title     = {Facial Expression Recognition using Convolutional Neural Networks:
               State of the Art},
  journal   = {CoRR},
  volume    = {abs/1612.02903},
  year      = {2016},
  url       = {http://arxiv.org/abs/1612.02903},
  archivePrefix = {arXiv},
  eprint    = {1612.02903},
  timestamp = {Wed, 07 Jun 2017 14:41:18 +0200},
  biburl    = {https://dblp.org/rec/bib/journals/corr/PramerdorferK16},
  bibsource = {dblp computer science bibliography, https://dblp.org}
}
```

Trechos importantes do *paper* :

- Finally, we confirm empirically that overcoming one such bottleneck improves performance substantially, demonstrating that modern deep CNNs achieve competitive results without auxiliary data or face registration (Section IV). An ensemble of such CNNs obtains a FER2013 [3] test accuracy of 75.2%, outperforming existing CNN-based FER methods.

Contribuições

1. Não necessidade de dados extras ou marcos faciais para obtenção de dados competitivos
2. Modelos não profundos, mas FC grandes e com 2 ou 3 camadas

Acurácia de 75.2%

Pontos importantes

- Pré-processamento
    - Equalização de histograma
    - Standardization
- Data augmentation
    - Horizontal mirror
    - Random crops
- Busca exaustiva para encontrar o valor de Dropout
- Treinamento em duas etapas
    - Primeiro treinamento das CNN com FC
    - Segundo, treinamento da rede com nova MLP, deixando fixos os parâmetros da CNN
- Modelos baseados em VGG, Inception e ResNet

---

3. Deep Learning using Support Vector Machines(Vencedor da competição do Kaggle)

```
@article{DBLP:journals/corr/Tang13,
  author    = {Yichuan Tang},
  title     = {Deep Learning using Support Vector Machines},
  journal   = {CoRR},
  volume    = {abs/1306.0239},
  year      = {2013},
  url       = {http://arxiv.org/abs/1306.0239},
  archivePrefix = {arXiv},
  eprint    = {1306.0239},
  timestamp = {Wed, 07 Jun 2017 14:41:28 +0200},
  biburl    = {https://dblp.org/rec/bib/journals/corr/Tang13},
  bibsource = {dblp computer science bibliography, https://dblp.org}
}
```

Trechos importantes do *paper*:

- Sem Trecho Importante

Contribuições

1. No lugar de softmax e FC utiliza SVM Linear

Acurácia de 71.2%

Pontos importantes

- Camadas Convolucionais utilizam ReLu, e seguem mesma estrutura dos meus modelos de CNN

- Modelo **sem** Ensemble

---

4. Fusing Aligned and Non-aligned Face Information for Automatic Affect Recognition in the Wild: A Deep Learning Approach

```
@article{Kim2016FusingAA,
  title={Fusing Aligned and Non-aligned Face Information for Automatic Affect Recognition in the
  author={Bo-Kyeong Kim and Suh-Yeon Dong and Jihyeon Roh and Geon-min Kim and Soo-Young Lee},
  journal={2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)},
  year={2016},
  pages={1499-1508}
}
```

Trechos importantes do *paper*:

- Trecho

Contribuições

1. Mostrou a importância de alinhamento da face para resultados competitivos com o estado da arte

Acurácia de 73.73%

Pontos importantes

- Alinha as imagens que são possível de ser alinhadas
- Classifica as expressões de acordo com a média simples das repostas das redes DCN e MLP
- Para as imagens que são alinháveis, são combinados as imagens alinhada e desalinhado
- Alinhamento é feito de maneira automática por modelos AMN(Alignent-Mapping network)
- As imagens que foram alinháveis obtiveram melhores resultados na classficação das expressões