EasyChair Preprint

№ 9929

Toward an Efficient Emotion Recognition from Facial Expressions Using ML

Hmad Zennou, Mohamed Ouhda and Mohamed Baslam

June 21, 2023

Toward an Efficient Emotion Recognition from Facial Expressions

Abstract:

This chapter discusses the abstraction and the process of recognition. It is concerned with the successive stages of processing that are involved in the encoding of simple stimuli and with the record that each stage produces. These issues lie within the areas of perception and memory, respectively. Because the recognition of stimuli is impossible without stored information, it will also be necessary to consider learning of trace systems applicable to the classification of patterns never before seen. The process of moving from the top to the bottom may be called abstraction. In psychological research, the term abstraction has been used in two different ways. One sense of abstraction involves the selection of certain portions or aspects of an experience. A second sense refers to the classification of a stimulus into a wider or more inclusive superordinate category. The second sense of abstraction has been used primarily with the investigation of object names. This sense of abstraction does not involve selection of any physical aspect of the stimulus, but rather a relationship between a particular stimulus name and another broader category name.
personally know many musicians who have made a living slinging code. I've gone to a few tech conferences that hosted evening festivities in which attendees gathered to jam, using instruments provided by conference sponsors. In fact, I once aspired to be a working musician, only to meander into programming later in life.

1. Introduction:

An introduction is the first paragraph of your paper. The goal of your introduction is to let your reader know the topic of the paper and what points will be made about the topic. The thesis statement that is included in the introduction tells your reader the specific purpose or main argument of your paper. These can be achieved by taking your introduction from "general" to "specific."

Think of an introduction paragraph in an academic paper as an upside-down triangle, with the broadest part on top and the sharpest point at the bottom. It should begin by providing your reader a general understanding of the overall topic. The middle of the introduction should narrow down the topic so your reader understands the relevance of the topic and what you plan to accomplish in your paper. Finally, direct your reader to your main point by stating your thesis clearly.

Introduction paragraph upside-down triangle
By moving from general subject to specific thesis, your audience will have a more concrete understanding of what your paper will focus on.

General

This refers to the broader topic you will address in your paper and its significance for the reader. For example, it might let your reader know you are writing about "climate change." Example: Climate change caused by humans is having a drastic effect on the world.

Narrowing

This is where you guide your reader to see your purpose for this particular paper. These sentences should give the reader an idea of what the context is for the topic. For example, it's not that you want to merely discuss climate change in general, but instead want to discuss the effects on yearly temperatures and how citizens can act. Example: However, the damage is not only affecting glaciers and rivers. Temperatures are starting to noticeable shift in cities and neighborhoods that have been otherwise consistent for centuries. Addressing the issue may require challenging decisions by individuals who have grown comfortable with their lifestyles and my be unaware of how their choices contribute to climate change.

Specific

This is where you narrow the focus to your argument, or your Thesis Statement. It is no longer about "climate change" or "human action," for example, but taking the argument all the way to your specific point. Example: While it has long been convenient to ignore how small changes may have a compounding effect on slowing climate change, it is vital to consider the extent to which measures such as eliminating single-use plastics can provide meaningful help.

gestures, voice, and physiological signals make up the multi-modality. Next, they propose a Convolutional Deep Belief Model (CDBN) for emotion recognition using this dataset [18]. Restricted Boltzmann Machines (RBMs) are an extension of Convolutional Restricted Boltzmann Machines (CRBMs). The RBMs are stacked to produce a convolutional deep belief network (CDBN). Layered generative models, or CDBNs, are generative models that are trained layer by layer. Ruiz-Garcia et al. [19] present a pre-trained deep CNN as a Stacked Convolutional AutoEncoder (SCAE). The SCAE is unsupervisedly taught in a greedy layer-wise manner. The model is trained using the Karolinska Directed Emotional Faces (KDEF) dataset [1] for face expression recognition.

The goal of this paper is to compare computer vision techniques for recognizing emotions in face expressions from image sequences. Models for static photos and image sequences are included in the datasets for comparison. It also applies to deep learning models with various inputs. The goal of this comparison is to determine the benefits and drawbacks of the various deep learning models that have been examined.

Paper organization:

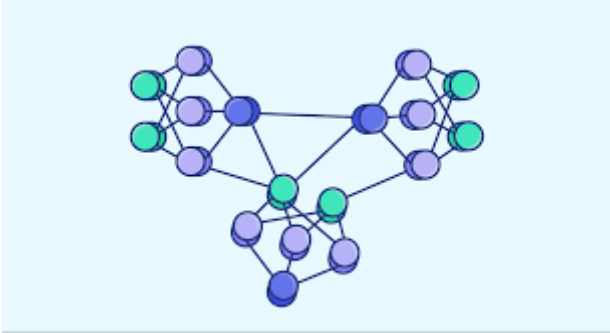
The rest of this paper is organized as follows; Section 2 presents the used deep learning models. Section 3 is dedicated to the experiments results and discussions, and finally conclusions are drown in Section 4.

2. Preliminaries :

This section introduces the deep learning models that will be used for testing. The section is split into three parts. The CNN-based models are presented first, followed by the 3D CNN model, and finally the RNN models.

1. recurrent neural network:

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes can create a cycle, allowing output from some nodes to affect subsequent input to the same nodes. This allows it to exhibit temporal dynamic behavior. Derived from feedforward neural networks, RNNs can use their internal state (memory) to process variable length sequences of inputs.[1][2][3] This makes them applicable to tasks such as unsegmented, connected handwriting recognition [4] or speech recognition.[5][6] Recurrent neural networks are theoretically Turing complete and can run arbitrary programs to process arbitrary sequences of inputs.[7]



This table presents the RNN parameters:

Table 1: CNN parameters

Parameters	values
Function	Sparse Categorical
Optimate	Adamsmit

2. 3D Convolutional Network:

Convolutional Networks (ConvNets) are a class of efficient neural networks that achieve impressive performances in perceptual tasks such as object recognition. Their architecture is loosely inspired by the visual cortex. In 2012 AlexNet, a type of ConvNet, won by a large margin the ILSVRC 2012 competition, starting the huge wave of interest in deep learning that continues today. In 2019, the state of the art architecture for object detection is ResNet, which is a type of ConvNet.

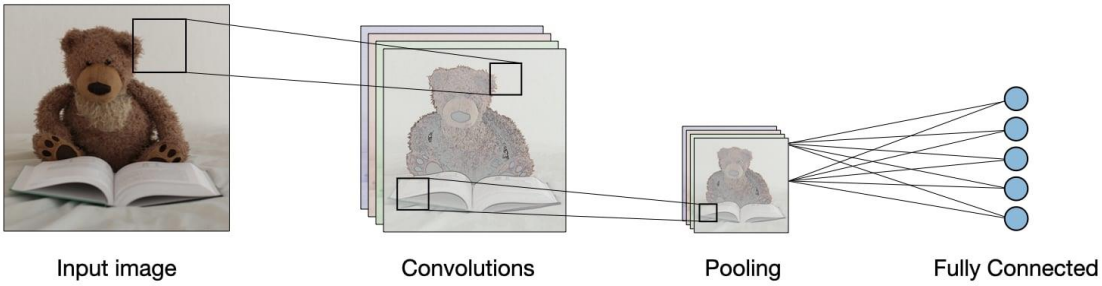


Figure 1:C3D Architecture (from [21])

The following table presents the C3D parameters:

Table 2: 3D CNN parameters

Parameters	values
Loss	Sparse Crossentropy
Optimize	Adamsat

Les LSTM (Long-Short-Term-Memory) et les GRU (Gated Recurrent Unit) se composent de plusieurs portes (respectivement 3 et 2) qui permettent d'oublier ou de mémoriser sélectivement les informations de la séquence temporelle précédente dans une mémoire dynamique.

Comme pour les réseaux de neurones traditionnels, les réseaux de neurones récurrents peuvent contenir plusieurs couches, ce qui leur permet de capturer davantage de non-linéarité parmi les données, mais augmente également le temps de calcul en phase d'apprentissage. On peut également combiner des couches récurrentes avec des couches classiques, telles que des couches denses (MLP) ou des couches de convolution (CNN). Dans la bibliothèque Keras, il existe 7 types de réseaux RNN : LSTM layer, GRU layer, SimpleRNN layer, TimeDistributed layer, Bidirectional layer, ConvLSTM2D layer (pour le traitement de vidéo) et le Base RNN layer.

Table 3: CNN + RNN parameters

Parameters	Values
Function	Categorical Crossentropy
Optimizer	Adam3
Layers	5

Datasets:

1) Garbage In, Garbage Out

Assembling and integrating training data comes with its own set of challenges. Web-sourced data can be inconsistent, irrelevant, low-quality, and uniquely formatted, creating problems as the data is streamlined for ingestion by model training systems. And if the data isn't coherent, the model will follow suit.

A model trained on irrelevant data decelerates go-to-market and wastes crucial development time when the results are less powerful and less accurate.

Clean, Relevant, Compliant

For over 20 years, SocialGist has provided clean and compliant data to enterprise companies and analysts. We're uniquely positioned to deliver the data developers need to train the next generation of machine learning models.

Sourcing training data from SocialGist guarantees:

Relevance: Training data should be closely related to the problem the model is trying to solve. For example, if the model is built to understand customer sentiment on products and services, relevant review data should be used for training.

Cleanliness: High-performing models require data that is free of noise, inconsistencies, and irrelevant information. The data also needs to be easy to integrate into the application. Clean data helps a model learn more efficiently, thus speeding development.

Maintenance: Training data needs to be continually updated and maintained.

With SocialGist, customers can focus on what they do best: creating cutting-edge ML solutions. We see the demand for quality data expanding as more players enter this space, and

our platform is prepared to meet this demand and help the trailblazers create the next wave of solutions that shape the world.
optimize their customer experience.

These same companies are taking note of how the recent advances in generally available artificial intelligence tools like ChatGPT will support everything from copywriting to product development.

Behind ML / AI is a robust dataset that has trained that machine to think, respond – even ideate – like a human. We're seeing the next wave of innovators entering ML / AI with the energy – and now the resources – to make big things happen. Digital data is the treasure trove for training the next wave of AI and machine learning models. But they're in for a rude awakening.

1. Our proposed method:

I understand the reservations that exist about the present method for choosing the President of the Commission, but the Treaties have been complied with.
The present method of calculation is clearly not altogether realistic.
We therefore have to recognise that the present method is no longer working.
We have exhausted the present method of the intergovernmental conference.
We must find a replacement for the present method of funding, which involves contributions from the national exchequers.
It therefore proposes to improve the present method of handling issues related to application and enforcement of Community law.
The present method can be used for all fertilisers mentioned in Annex I which contain exclusively nitric, ammoniacal or urea nitrogen.
the Commission should present the method for applying 1/6ths of 2007 appropriations' 'recommitments' to be spread over the years 2008-2013.
Firstly, under the present system the reference method for the European Union is the ISO method.
The Århus Convention is undoubtedly a real advance, as it plays a part in establishing what has become essential transparency, and we are called upon here to present methods of providing the public with access to justice.
The present method applies to pure substances that do not dissociate or associate and that do not display significant interfacial activity.
Thus, the numerous phenomena and mechanisms involved in the process of adsorption of a chemical by soil cannot be completely defined by a simplified laboratory model such as the present method.
Contrast that with this present method: one far-reaching proposal for a directive literally dumped on the table at the end of the last mandate.

At present such methods utilise high-resolution gas chromatography/high resolution mass spectrometry (HRGC/HRMS).
France has requested the Commission to authorise the replacement of the formula used in the 'CSB Image-Meater' method for grading pig carcasses on its territory as the present grading method needed technical adaptation.
Finally, I should like to make it clear that it is particularly important to harmonize the methods of sampling and monitoring lead concentrations; because the fact is that at present the methods for monitoring the concentration of lead in water are not harmonized.
One would almost be inclined to think that none of the commercial banks stand to benefit from modernising their present working methods.
The present centralised method, which means that there is a backlog of cases, is inevitably cannot guarantee legal certainty.
Firstly, the net present value method is the commonly used methodology in the energy sector as well as in other industrial sectors.
Today I'll present an alternative method to a posterior hip dislocation.
Some examples from the web:
We have a resolution dated 19 November 1997 in which we propose a method for the reform of the institutions.
The Commission should also propose a method for monitoring the long-term success of projects, in particular in the priority area Nature and Biodiversity.
I also propose a method of improving the competitiveness of travel agencies when selling tourism products within the Union to residents of third countries.
To avoid these difficulties, the Commission proposes a method that:
The French Presidency will propose a method and, I hope, a solution in agreement with the Irish Government, either in October or December.
In this Directive the Commission proposes a new method for establishing the authorised level of nitrates/nitrites in food.
A new proposal to amend the so-called Eurovignette directive proposes a harmonised method for charging full infrastructure costs to heavy goods vehicles.
I believe that the Research Committee has done its work on the fifth framework programme; it is now up to Mr Kinnock to propose a method of funding the infrastructure.
The Commission noted that the beneficiaries receive an indirect advantage and asked the interested parties to propose methods for accurately quantifying this advantage.
In addition, the ECB and Eurostat published a report on foreign direct investment, which proposes improved methods for compiling such statistics.
If methods listed in this Regulation are not suitable, the applicant shall provide adequate reasoning and propose a new method.
The Danish authorities propose two methods for assessing the productivity of DSB's activities:
Furthermore, this study proposes two methods for the ex ante assessment of the positive effects of marketing services agreements: a 'cash flow' approach and a 'capitalisation' approach.
Many colleagues propose alternative research methods - for example, that we should use adult stem cells.
I agree with the report's call to take action and propose methods of harmonising contract law practice at EU level, which would ensure equal and fair conditions for market participants.
It is a very important step forward that the Commission proposes a harmonised allocation method.
The Commission proposes a streamlined Open Method of Coordination (OMC) for social protection and social inclusion, in line with its first plans from 2003.

On 1 March 2011, the Belgian authorities sent the Commission a report prepared by Charles River Associates (CRA), which critically evaluates the WIK study and proposes an alternative method for benchmarking the reasonable profit of DPLP.
The applicant should also propose a validated method of sampling and detection for the primary products to be used for control of compliance with the provisions of this Regulation. You rapporteur proposes adopting the method of the Convention.

2. Implementation and discussion:

The steps involved in pre-processing are discussed in this section. The SASE-FE and OULUCASIA datasets both go through the same steps. Videos cannot be used as inputs to non-temporal models; instead, frames from videos must be extracted and used as inputs to the models. A vector containing a succession of these frames is fed into the temporal models as an input.

- Pre-processing :

Each frame of the videos was extracted as an image using a pre-processing technique. There are some frames that are useless because the videos begin with a neutral face expression and then the participant makes the facial expression that corresponds to the emotion. This is considered by the pre-processing, which only keeps frames from half of the video duration to 80% of the video duration.
This guarantees that the frames obtained convey the desired emotion.
Hassner et al [5] proposed a procedure called frontalization, which was used to perform a second change on the datasets. By transforming unconstrained perspectives to constrained, forward facing faces, this method rotates and scales the participant's face, limiting the variability of the location of the faces. Although frontalization can assist reduce variability, it also has significant disadvantages, particularly when the face is partially occluded.
Hassner also proposes soft symmetry, which allows for the estimation of occluded sections of the face when both parts of the face differ. Blending techniques are used to create a symmetrical image on both sides. Figure 5 exhibits a symmetrical image and a symmetrical image with soft symmetry.

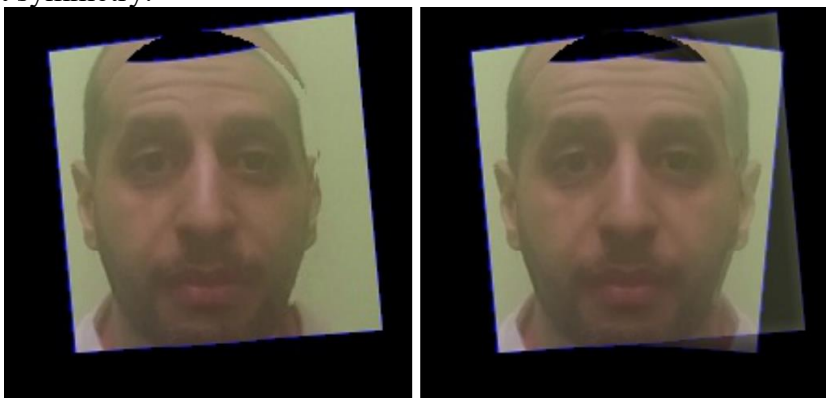


Figure 2: Left image presents a Soft Symmetry frontalization process. Right image corresponds to an image with No Symmetry frontalization.

Figure 5 shows the Hassner method [5] for obtaining face landmarks, which consists of 68 fiducial points. These characteristics correspond to sites in the mouth, nose, and eyes, among other places.
These landmarks are then fed into the VGG-Face model as a second input. Because the VGG-Face only accepts photos as input, an intermediate fusion approach is necessary in the first fully connected layers.

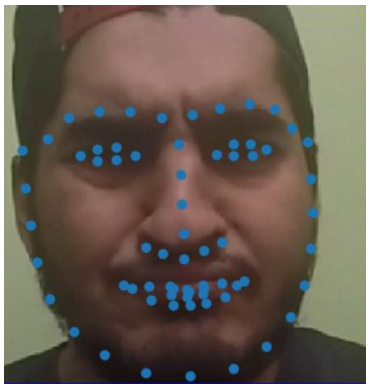


Figure 3: 68 fiducial points superimposed on the detected face

- Experimental Results

This section summarizes the findings from many experiments conducted on various dataset configurations with various model layouts. Only the test accuracy is shown for general purposes.

A. SASE-FE Emotion Results

The first series of experiments focused solely on classifying the six emotions, combining both actual and false feelings. Using the SASE-FE dataset, these experiments include fine-tuning the model. The goal is to evaluate various configurations and select the model with the best test accuracy.

- No Pre-Processing

The experiments are designed to see if adding three fully connected layers and/or a pooling layer after the convolutional layers enhances the CNN's performance. Other tests include freezing all the convolutional layers, freezing only the first few layers, and not freezing any of the layers at all.

The results are shown in the next table.

Table 4: Emotion No Pre-processing configurations Accuracy

Used configuration	Accuracy
No Freezed + No Pooling	0.2842
First Freezed + 3 Fully Connected + No Pooling	0.1970
All Freezed + 3 Fully Connected	0.4281
All Freezed + 3 Fully Connected + Average Pooling	0.4207
All Freezed + 3 Fully Connected + Max Pooling	0.4375
All Freezed + No Pooling	0.4250
All Freezed + Avg Pooling	0.4226
All Freezed + Max Pooling	0.4310

With a test accuracy of 0.4375, the optimal configuration is all Convolutional layers freeze with three Fully Connected layers at the end and Max Pooling layer at the end. Only this configuration will be used in the next experiments.

- Frontalization

The next experiment is to apply the frontalization pre-processed dataset after obtaining the optimal architecture from the previous tests. Both soft and no symmetry are visible in the experiments. The results are shown below.

Table 5: Emotion Frontalization configurations Accuracy

Used configuration	Accuracy
Soft Symmetry	0.454692
No Symmetry	0.594999

The difference between no symmetry and soft symmetry is enormous, as shown in the previous table. The difference amounts to over 15%.

Hassner et al. [6] noted how soft symmetry "may actually be unneeded and potentially even harmful; harming rather than boosting face recognition ability" in some circumstances. Soft Symmetry blends the identified facial features with the surface by modifying it. This mixing, however, is an approximation that can cause noise. Looking at the accuracies, it appears that performing soft symmetry has a significant impact on emotion identification. As a result, the next experiment will solely use no symmetry.

- Two-Stream CNN

According to the previous section's study, no symmetry leads to greater accuracy. Nonetheless, it will be fascinating to see if merging the no symmetry dataset with the extracted face may help boost the accuracy even further. The following experiments include employing a two-stream CNN, which combines a CNN with no symmetry dataset as input and another CNN with no pre-processing dataset as input. Both CNNs are fused before the fully connected layers to achieve this. The architecture is unchanged after the fusing layer. The results are presented below:

Table 6: Emotion Fuse-Stream configuration Accuracy

Used Configuration	Accuracy
Two-Stream	0.583234

The test accuracy for the two-stream CNN was 0.5832, which is lower than the 0.5949 for the No Symmetry CNN. The empirical evaluation with this dataset demonstrates that utilizing Soft-Symmetry is not useful to the emotion recognition task in the situations investigated here. Introducing Soft-Symmetry, as noted by Hassner et al. [12], may "create issues whenever one side of the face is obscured... rendering the final product unrecognizable." Finally, Soft-Symmetry photos will not be used in future models.

B. SASE-FE Hidden Emotion Results

The dataset was divided into fake and real emotions in this set of experiments, yielding a total of 12 classes. Each of the six emotions is divided into two categories: fake and real. Because there are now 12 courses, the test accuracy is projected to be substantially lower.

- Still Images Input

The VGG-Face is used in the first set of trials, with one experiment using photos of the face and the other using the frontalized face. Table 7.4 reveals that frontalization data has a modest advantage over only the face in terms of accuracy.

The second experiment is a two-stream CNN that employs both datasets, with one stream using frontalization and the other using face frontalization. One CNN improves the accuracy of the combined CNN significantly. There are almost four decimal points in the increase.

The third experiment is a middle fuse CNN that uses frontalization as an input and geometry data as output. The test accuracy of 0.2994 improves when the geometry is added to the base accuracy.

Table 7: CNNs Hidden configurations Accuracy

Model	Used configuration	Accuracy
CNN	Face Frontalization	0.2806 0.2866
Two-Stream CNN	Frontalization + Face	0.3206
CNN + Geometry	Frontalization + Geometry	0.2994

- Image Sequences Input

The next set of experiments differs from the prior ones in that they now include temporal data. These tests are fed a 5-frame vector containing the range of articulations that each participant uses to express the emotion.

A 3D CNN is used in the first experiment of this type. To get decent performance, most 3D CNNs require a large amount of data. This performance issue is well-known, but the experiment aims to investigate if a 3D CNN can train even with a short dataset. The accuracy of the exam was 0.1281 percent. Despite the decreased accuracy, it is crucial to note that this model incorporates temporal information from image sequences. The model does not employ any fine-tuning and starts the learning process from the beginning.

The fine-tuned model developed with the frontalization preprocess provided in table 7.5 is used in the next tests. Two models were trained, one of which was fine-tuned with a 5-frame input vector. The second step is to extract features from a pre-trained CNN with a vector size of 4096; PCA is then applied to the feature vector to minimize its size, with just the first 100 eigenvectors used.

The second experiment combines a CNN with an LSTM. It's fascinating to note that feature vectors outperform picture vectors in the results. The conclusion is that PCA aids in obtaining the most variables, while the LSTM learns the differences that distinguish each emotion.

The final experiment is the CNN, but this time with a GRU on top. Surprisingly, the model with image vectors as input performs horribly. One thing to keep in mind is that SASE-FE frames begin with a neutral face. The 5 frames picture vector is relatively modest to display the emotion's complete spectrum of expression. The features vector model, on the other hand, has a very high accuracy. Higher even than the LSTM model.

Table 8: Temporal Hidden configurations Accuracy

Model	Used configuration	Accuracy
3D CNN	Frontalization	0.128125
CNN + LSTM	Features Image	0.159200 0.148684
CNN + GRU	Features Image	0.183311 0.084134

C. OULU-CASIA Results

The following experiments were conducted specifically to classify the OULUCASIA dataset's six emotions.

- Still Images Input

The first set of experiments involves fine-tuning the VGG-Face using photos that have not been pre-processed; the second set of tests involves frontalization pre-processed images. Preprocessed photos outperformed non-processed images by nearly 0.03 percent.
The accuracy of the models is improved by this pre-processing. This is because frontalization normalizes the faces and aids the model in learning the differences between emotions; normalizing the images reduces noise, therefore the performance improves in this situation.

The second experiment employs a two-Stream CNN to investigate if combining both the face and frontalization datasets improves the accuracy of the prior two. Frontalization received a 0.2659 percent in earlier studies. With a higher score of 0.2737 percent, the two-stream CNN outperforms the model. This ensures that the model learns all feasible information from the face, even information that may have been lost during frontalization pre-processing.

The third experiment employs an intermediate fusion technique, employing one CNN before concatenating data from the face's geometry after the final convolutional layer. The geometry is made up of 68 fiducial points that are normalized in a new center. It's fascinating to note that this model beats all previous studies by a significant margin, with a test accuracy of 0.4411 percent. The increase is 0.17 percent above the previous highest model.
The following table presents the results.

Table 9: Temporal configurations Accuracy

Model	Configuration	Accuracy
CNN	Face Frontalization	0.2386 0.2659
Two-Stream CNN	Face + Frontalization	0.2737
CNN + Geometry	Frontalization + Geometry	0.4411

- Image Sequence Input

The next experiments differ from the previous ones in that they now use temporal data from image sequences. In this situation, the input is a frame sequence.
Each frame is made up of five consecutive photos taken from the videos.
A 3D CNN is used in the first experiment. Although 3D CNNs are known to require a lot of data to learn well, this model's performance is comparable to the other temporal models shown below.
The fine-tuned model developed with the frontalization preprocess provided in table 11 is used in the following tests. Two types of inputs are tested: image vectors and features vectors. They all follow the same steps.
In the second experiment, a CNN is combined with an LSTM. The picture vector input is noticeably more accurate than the feature vector in this scenario. The difference is less than 0.02%.
A CNN with a GRU built on top is the third and last experiment. The image vector has a greater test accuracy than the features vector, as with the CNN+LSTM. The GRU features

vector, on the other hand, has a larger disparity between features and image vectors, at roughly 0.05 percent.

Table 12: Accuracy of Temporal Test

Model	Configuration	Accuracy
3D CNN	Frontalization	0.2000
CNN + LSTM	Features	0.2062
	Image	0.2209
CNN + GRU	Features	0.1797
	Image	0.2241

D. Discussion

For both datasets in table 13, this section discusses the best model for each test. The VGG-Face and C3D models were investigated in this work. With multi-modality, recurrent neural networks, the CNN-based models is improved even further. A voting mechanism is considered when evaluating the models.

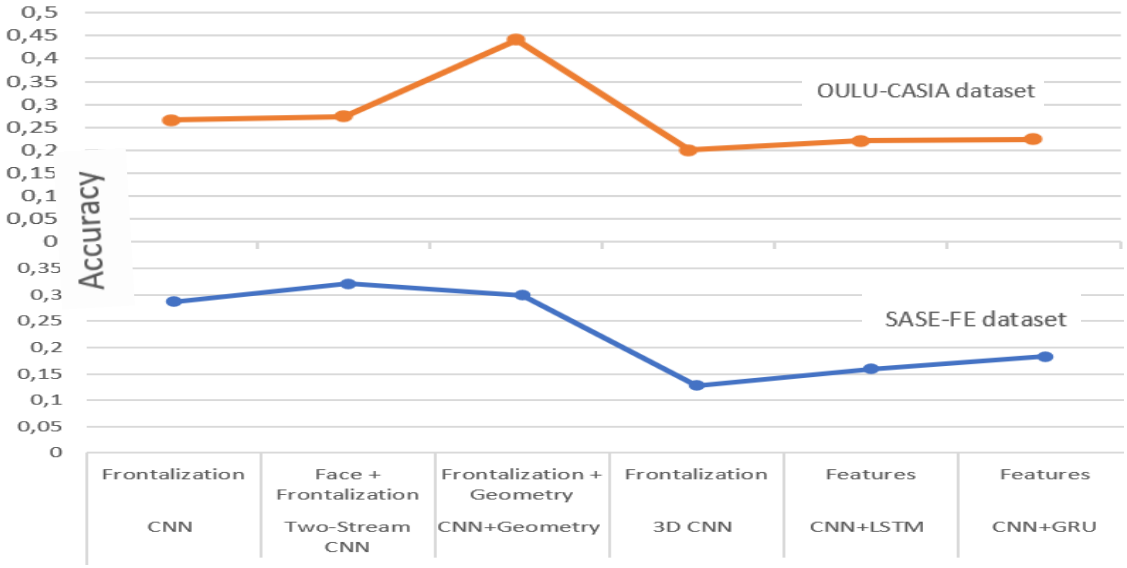
The basic CNN is used in the first experiment with pre-processing and no pre-processing data; in both datasets, the frontalization method has a higher accuracy. Frontalization is the process of mapping the face in a confined, forward-facing position. This reduces the variability in face location in the dataset, allowing the models to focus solely on learning the variability of emotion recognition.

In both datasets, the accuracy of the 2-Stream CNN is higher. However, it is the best image model in SASE. Overall, including both inputs aid the models in learning the difference that may have been lost during the frontalization process, which might result in face misalignment.

Table 13: Summary of used models' accuracy

Dataset	Model	Best used Configuration	Accuracy
SASE-FE dataset	CNN	Frontalization	0.2866
	Two-Stream CNN	Face + Frontalization	**0.3206**
	CNN+Geometry	Frontalization + Geometry	0.2994
	3D CNN	Frontalization	0.1281
	CNN+LSTM	Features	0.1592
	CNN+GRU	Features	**0.1833**
OULU-CASIA dataset	CNN	Frontalization	0.2659
	Two-Stream CNN	Face + Frontalization	0.2737
	CNN+Geometry	Frontalization + Geometry	**0.4411**
	3D CNN	Frontalization	0.2000
	CNN+LSTM	Image	0.2209
	CNN+GRU	Image	**0.2241**

Summary of used models' accuracy

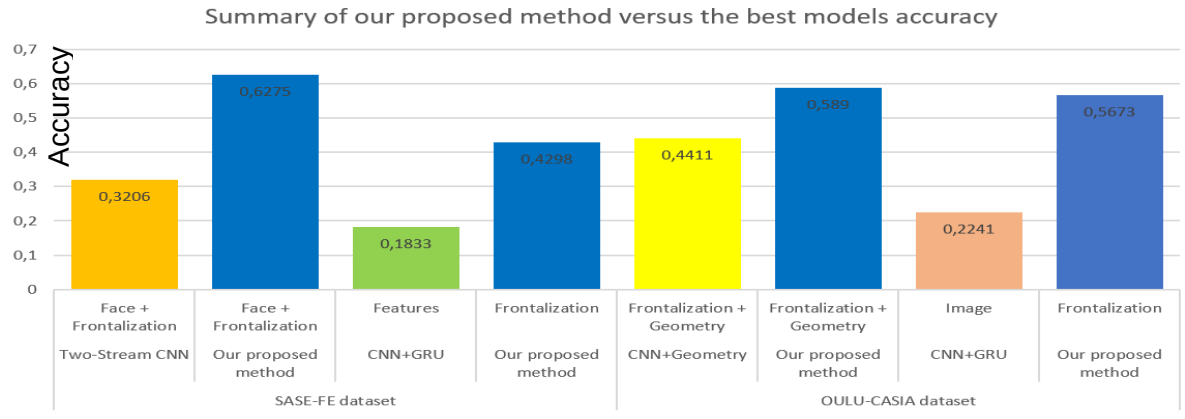


The geometry data improves the performance of middle fusion models in both the SASE-FE and OULU-CASIA datasets when compared to just one input. However, the OULU-CASIA boost is substantially bigger than the SASE-FE boost. OULU earns a 0.17 percent rise compared to SASE's 0.1 percent. All the models evaluated have a higher boost than OULU's Geometry model.

Afterwards, we compare the best models with our proposed method. The results are shown below:

Table 14 :10 Summary of our proposed method versus the best models accuracy

Dataset	Model	Best used configuration	Accuracy
SASE-FE dataset	Two-Stream CNN	Face + Frontalization	0.3206
	Our proposed method	Face + Frontalization	**0.6275**
	CNN+GRU	Features	0.1833
	Our proposed method	Frontalization	**0.4298**
OULU-CASIA dataset	CNN+Geometry	Frontalization + Geometry	0.4411
	Our proposed method	Frontalization + Geometry	**0.5890**
	CNN+GRU	Image	0.2241
	Our proposed method	Frontalization	**0.5673**



The results above show that our method outperforms the best models in all configuration methods and in both datasets. For SASE-FE dataset, as we can see our method achieves an averaged accuracy of 62% in the first set of experiments and 42% in the second set. The same for OULU-CASIA dataset, which performs 58% in the first set and 56% in the second set. The results of our method greatly exceeded those of the other methods, 30% compared to Two-Stream CNN, and 24% compared to CNN+GRU for the SASE-FE dataset.

3. Conclusion

CONCLUSION DU RAPPORT DE STAGE
Je tire un bilan très positif de ce stage, qui fut une expérience très enrichissante tant sur le plan professionnelque personnel. Sur le plan professionnel d'abord, j'ai pu appréhender toutes les facettes du métier dePOSTEOCCUPÉ, notammentLISTER LES MISSIONS / TÂCHES RÉALISÉES. J'ai donc rempli les objectifs fixés, à sa-voir :LISTER LES OBJECTIFS DONNÉS PAR L'ENTREPRISE. Sur le plan personnel ensuite, j'ai pu comprendrequeLISTER LES DIMENSIONS DU POSTE QUE VOUS AVEZ LE MOINS APPRÉCIÉ, ne représentait pas ce quime correspondait le plus. Au cours de cette période, comme dans toute phase d'apprentissage, il m'est parailleurs arrivé de faire quelques erreurs comme :LISTER VOS ERREURS, j'ai pu rapidement les corriger enDÉMONTRER COMMENT VOUS LES AVEZ CORRIGÉES.Grâce aux acquis d'une méthodologie de travail forte que l'entrepriseNOM DE L'ENTREPRISE m'a transmise,combinée à la formation théorique que j'ai reçue, je suis aujourd'hui en mesure d'affirmer qu'à la question :"PROBLÉMATIQUE", il y a plusieurs éléments de réponses, à savoir :RÉPONS

Bibliographie

[1] M. G. Calvo and D. Lundqvist. Facial expressions of emotion (KDEF): Identification under different display-duration conditions. Behavior Research Methods, 2008.
[2] F. Chollet et al. Keras, 2015.
[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 2005.
[4] C. Feichtenhofer, A. Pinz, and A. Zisserman. Convolutional Two-Stream Network Fusion for Video Action Recognition. 2016.
[5] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective Face Frontalization in Unconstrained Images.
[6] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015.
[7] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.
[8] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional Architecture for Fast Feature Embedding.
[9] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive Database for Facial Expression Analysis.
[10] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale
Video Classication with Convolutional Neural Networks. In Large-scale Video Classication with Convolutional Neural Networks, 2014.

[11] P. Liu and L. Yin. Spontaneous Thermal Facial Expression Analysis Based On TrajectoryPooled Fisher Vector Descriptor.
[12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression.
[13] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding Facial Expressions with Gabor Wavelets.
[14] N. Neverova, C. Wolf, G. Taylor, and F. Nebout. ModDrop: Adaptive multi-modal gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.
[15] I. Ofodile, K. Kulkarni, C. A. Corneanu, S. Escalera, X. Baro, S. Hyniewska, J. Allik, and G. Anbarjafari. Automatic Recognition of Deceptive Facial Expressions of Emotion. 2017.
[16] M. Osadchy, Le Cun. Yann, and M. L. Miller. Synergistic Face Detection and Pose Estimation with Energy-Based Models. The Journal of Machine Learning Research, 2007.
[17] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep Face Recognition. In Procedings of the British Machine Vision Conference 2015, 2015.
[18] H. Ranganathan, S. Chakraborty, and S. Panchanathan. Multimodal Emotion Recognition using Deep Learning Architectures.
[19] A. Ruiz-Garcia, M. Elshaw, A. Altahhan, and V. Palade. Stacked Deep Convolutional Auto-Encoders for Emotion Recognition from Facial Expressions.
[20] N. Sarode and S. Bhatia. Facial Expression Recognition. International Journal on Computer Science and Engineering, 02(05):1552–1557, 2010.
[21] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3D convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, 2015.
[22] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri. Learning spatiotemporal features with 3D convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, 2015.
[23] H. C. Vijay Lakshmi and S. PatilKulakarni. Segmentation Algorithm for Multiple Face Detection in Color Images with Skin Tone Regions using Color Spaces and Edge Detection Techniques. IEEE Signal Acquisition and Processing, 2010. ICSAP'10. International Conference on, pages 162–166, 2010.
[24] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.
[25] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang. A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Transactions on Multimedia, 2010.
[26] L. Wolf, T. Hassner, and I. Maoz. Face Recognition in Unconstrained Videos with Matched Background Similarity.
[27] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietik¨ainen. Facial expression recognition from near-infrared videos. Image and Vision Computing, 2011.
[28] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4489-4497
[29] He, K., Zhang, X., Ren, S., Sun, J. (2014). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8691. Springer, Cham.

[30] Zhao, X. et al. (2016). Peak-Piloted Deep Network for Facial Expression Recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9906. Springer, Cham.
[31] Mengyi Liu, Ruiping Wang, Shaoxin Li, Shiguang Shan, Zhiwu Huang, and Xilin Chen. 2014. Combining Multiple Kernel Methods on Riemannian Manifold for Emotion Recognition in the Wild. In Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14). Association for Computing Machinery, New York, NY, USA, 494–501.