

Réseaux de neurones et systèmes flous
SYS843

Professeur responsable
Éric Granger
Ismail Ben Ayed

Reconnaissance et classification d'émotions
Étude expérimentale

Auteur
Vincent Lynch – LYNV24059408

École de technologie supérieure
22 avril 2018

Table des matières

1 – Introduction	4
1.1 – Domaine d’application	4
1.2 – Problématique.....	4
1.2.1 – Expression spontanée ou délibérée.....	4
1.2.2 - Encodage et caractéristiques faciales.....	4
1.2.3 - Obstruction et angle du visage	5
1.2.4 - Éclairage et résolution	5
1.3 – Objectif du projet.....	5
1.4 – Structure du document.....	5
2 - Sommaire des techniques étudiées.....	7
2.1 - Présentation des systèmes	7
2.1.1 – Prétraitement de la base de données.....	7
2.1.2 – Compression de la base données	8
2.1.2 – Entraînement et classification	9
2.2 - Description algorithmique	11
3 - Méthodologie expérimentale	12
3.1 - Base de données	12
3.2 - Protocole d’évaluation.....	13
3.3 - Mesures de performances.....	13
4 - Résultats de simulation.....	14
4.1 - Présentation des résultats	14
4.2 - Interprétation des résultats.....	16
5 - Conclusion.....	18
5.1 - Recommandations	18
Bibliographie	20

Liste des figures

Figure 1 – Extraction du visage utilisé pour les 2 expérimentations.....	7
Figure 2 – LBP utilisé pour la 1 ^{ère} expérimentation	8
Figure 3 – Première méthode de tri de la base de données	12
Figure 4 - Deuxième méthode de tri de la base de données.....	13
Figure 5 – Matrice de confusion de la première expérimentation	15
Figure 6 – Matrice de confusion de la deuxième expérimentation	15
Figure 7 – Précision du deuxième model en fonction des époques.....	16

Liste des tableaux

Tableau 1 – Architecture de l’objet contenant les informations de la base de données	9
Tableau 2 – Architecture du CNN utilisé pour la 2 ^e expérimentation	10
Tableau 3 – Résumé du nombre d’images traitées selon la technique utilisée	12
Tableau 4 – Synthèse de la précision des expérimentations.....	14
Tableau 5 – Synthèse des temps de calcul des deux expérimentations	14

1 – Introduction

1.1 – Domaine d'application

La reconnaissance des multiples émotions qu'un être humain peut exprimer est un sujet étudié extensivement et bien documenté dans le domaine de la psychologie et bien d'autres. Plus récemment, avec la remontée en popularité de l'intelligence artificielle, beaucoup de chercheurs se sont penchés sur la reconnaissance des émotions chez l'être humain. Un humain utilise plusieurs mécanismes pour exprimer ses émotions comme, par exemple, le ton de voix ainsi que la communication non verbale. Plusieurs systèmes ont été conceptualisés basés sur ces mécanismes pour reconnaître les émotions d'un humain lorsque celui-ci interagit avec un système informatique.

Un système ayant la capacité de reconnaître les expressions faciales d'un être humain se verrait utilisé dans d'innombrables domaines d'applications. En contrepartie, pour qu'un tel système soit utilisable dans le monde réel, il est impératif que celui-ci aille la capacité de déterminer des émotions assez rapidement pour une utilisation en temps réel.

1.2 – Problématique

L'application désirée est extrêmement difficile à réaliser en grande partie dû aux variabilités qui sont présentes dans l'expression des émotions sur un visage humain. Non seulement il existe plusieurs émotions uniques qu'un humain peut exprimer, mais celles-ci sont souvent différenciées par de petites différences tel qu'un sourcil légèrement plus élevé que l'autre. Dans la section suivante, on présente les principales difficultés qui attendent tous chercheurs qui tentent d'entreprendre la reconnaissance des émotions par le visage.

1.2.1 – Expression spontanée ou délibérée

La grande majorité des bases de données sont recueillies en indiquant au sujet quelle émotion performer lors de la phase d'enregistrement. Les émotions exprimées consciemment et inconsciemment utilisent deux différents systèmes moteurs. Ainsi, il est fort probable qu'une machine ayant appris avec une base de données typiques ne puisse obtenir de bonne performance dans un environnement d'utilisation réaliste. À moins qu'une base de données contienne les deux types d'émotions, un système fiable ne pourra être bâti. De plus, les émotions spontanées sont typiquement moins accentuées et plus subtiles qu'une émotion délibérée. Les bornes décisionnelles devront sans aucun doute être plus sensibles sur un système voulant être utilisé dans une application réelle.

1.2.2 - Encodage et caractéristiques faciales

Les techniques d'encodage des différentes sections du visage d'un sujet est une sphère du domaine qui est ouverte aux suggestions. Malgré le fait qu'on ait développé plusieurs techniques pour se faire, les chercheurs sont encore loin d'avoir une solution générale. De plus, il n'est pas seulement question de représenter une section du visage dans un langage machine,

mais aussi de trouver quelles sections ou caractéristiques du visage représenteront le mieux notre problématique. Évidemment avec de nouvelles techniques tels que les CNN, il est possible de faire l'apprentissage de caractéristiques. Ainsi, le modèle décide par lui-même quelle information est la meilleure pour différencier les émotions. Malheureusement, comme nous apprendrons plus tard dans cette étude expérimentale, ces nouvelles techniques n'augmentent pas les performances de façon significative.

1.2.3 - Obstruction et angle du visage

Plusieurs bases de données utilisent une caméra frontale pour capturer le visage d'un sujet. Un système entraîné avec une telle base de données aura une grande difficulté à classer différentes émotions lorsque le sujet n'est pas parfaitement perpendiculaire à l'objectif de la caméra.

1.2.4 - Éclairage et résolution

Évidemment, si le visage d'un sujet est enregistré dans un environnement avec peu de lumière, la quantité d'informations contenues dans l'image sera réduite et influencera les résultats. L'éclairage pourrait aussi être simplement inégal, réduisant les informations contenues dans certaines sections du visage. De plus, la résolution d'une image influencera elle aussi la quantité d'informations extraites. Malgré que le visage soit généralement placé au centre de l'image dans un environnement contrôlé, le visage sera potentiellement petit relatif à l'objectif d'une caméra. En contrepartie, une haute résolution augmente les temps d'apprentissage et de calcul.

Les problématiques d'obstruction, d'éclairage, de résolution et d'angle du visage peuvent facilement être remédiées dans un environnement de test. Cependant, lorsqu'un système pré-entraîné sera utilisé dans un environnement plus réaliste, il sera inévitablement présenté à ces types de problèmes. Ainsi, il est impératif de construire des systèmes robustes capables de s'adapter aux différents types d'images.

1.3 – Objectif du projet

L'objectif ultime du projet serait de déterminer lequel des deux systèmes étudiés en profondeur dans la synthèse de littérature est le plus efficace pour la reconnaissance des émotions. Évidemment, pour y parvenir, nous devons concevoir et réaliser les deux systèmes et les comparer avec une base de données communes. En modifiant différents paramètres tels que le nombre de données d'apprentissage et la structure des réseaux, nous serons en mesure d'évaluer objectivement la performance de chaque système. Nous serons en mesure de déterminer si les techniques modernes de reconnaissance de formes sont en mesure de surpasser les techniques utilisées depuis plusieurs années

1.4 – Structure du document

À la suite de cette introduction, nous entamerons une description des deux différents systèmes conçus pour cette étude expérimentale. D'une première part, un modèle basé sur un classificateur à vecteur de support sera présenté suivi d'un modèle basé sur un classificateur à

réseau de neurones. Nous tenterons de présenter les techniques utilisées lors de la réalisation expérimentale. Une fois les deux différentes approches présentées, nous nous concentrerons sur l'évaluation de celles-ci. Nous tenterons de pousser l'analyse plus en profondeur pour comparer ultimement les deux approches et en retirer une conclusion. À la toute fin du document, quelques recommandations seront présentées pour donner une direction future à cette étude expérimentale.

2 - Sommaire des techniques étudiées

2.1 - Présentation des systèmes

Les deux différentes réalisations expérimentales ont été grandement inspirées des articles 1 et 2 présentés dans la revue de littérature. Essentiellement, les deux différentes approches ont été réalisées sur la même base de données pour une meilleure comparaison. À des fins de simplicité, la 1^{ère} expérimentation est reliée au 1^{er} article¹ et la 2^e expérimentation au 2^e article² de la revue de littérature.

2.1.1 – Prétraitement de la base de données

Le même prétraitement de la base de données est utilisé pour les deux systèmes. D'une première part, les visages présents dans chaque image de la base de données sont détectés à l'aide du module Python "openCV". La méthode utilisée pour détecter un visage est l'extraction de caractéristiques pseudo-Haar offerte par le module. Une fois le visage détecté, il suffit de le redimensionner à une résolution plus acceptable pour les traitements futurs. Pour les deux expériences, une dimension de 100x100 pixels a été choisie. Il est possible d'utiliser une autre dimension d'image mais, après une lecture extensive de la littérature, cette résolution semble typique et souvent utilisée. Une simple interpolation linéaire est utilisée pour réduire la dimension des images.



Figure 1 – Extraction du visage utilisée pour les 2 expérimentations

Une autre étape relativement simple du prétraitement consiste à copier les étiquettes de chaque image dans le même répertoire où son image correspondante est localisée. Cette étape simplifie les algorithmes d'apprentissage et est nécessaire puisque la base de données Cohn-Kanade

¹ Shan, Caifeng, et al. "Robust facial expression recognition using local binary patterns." IEEE International Conference on Image Processing 2005, 2005, doi:10.1109/icip.2005.1530069.

² Sang, Dinh Viet, and Nguyen Van Dat. "Facial expression recognition using deep convolutional neural networks." Knowledge and Systems Engineering (KSE), 2017 9th International Conference on. IEEE, 2017.

contient deux fichiers séparés pour les images et les étiquettes. Pour nos deux expérimentations, seulement les étiquettes et les images sont utilisées.

Base de données Cohn-Kanade

- └ extended-cohn-kanade-images
- └ Emotion_labels
- └ FACS_labels
- └ Landmarks

2.1.2 – Compression de la base données

Puisque les scripts Python de compression de données et d'apprentissage sont séparés, il est nécessaire de concevoir une méthode pour transférer les données d'un script à l'autre. Pour se faire, le tableau final où se trouve tous les vecteurs de données et leur étiquette correspondante utilise le module "pickle" qui permet ainsi la sérialisation des données.

1^{ère} expérimentation – LBP & SVM

Lorsque toutes les images originales sont remplacées par les visages redimensionnés, il est possible de passer à l'étape de compression des données. La première étape consiste à appliquer un algorithme de "local binary pattern" à l'ensemble du visage convertie en niveau de gris. Pour cette expérimentation, seulement les patrons uniformes ont été utilisés ce qui réduit la grandeur de l'histogramme de 255 à 59. Une fois la nouvelle image créée, il est possible de subdiviser un visage en 4x4. Pour chaque sous portion du visage, un histogramme des différentes valeurs de pixel est extrait. La concaténation des 16 histogrammes forment le vecteur de caractéristique qui sera présenté à l'entrée de notre classifieur.

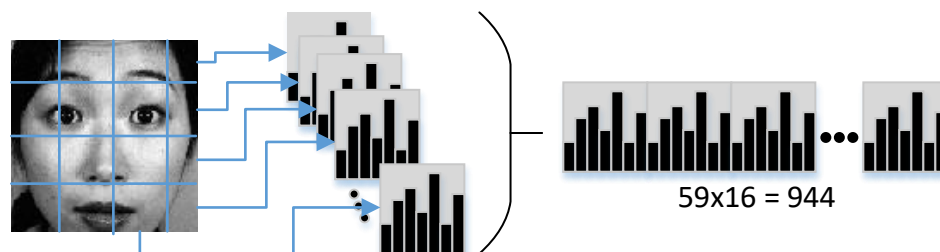


Figure 2 – LBP utilisé pour la 1^{ère} expérimentation

Évidemment, il est essentiel d'effectuer le traitement décrit ci-haut à toutes les images de la base de données. L'algorithme de compression crée alors un tableau contenant toutes les informations nécessaires pour l'entraînement et la phase de test. Le tableau ci-dessous présente la forme générale de l'ensemble de données final. Évidemment, "training data", "Validation data" et "test data" sont tous de grosseur différente; plus spécifiquement 80%, 10% et 10% respectivement.

Training data		Validation data		Test data	
Data	Label	Data	Label	Data	Label
[0.2, ..., 0.1]	[5]	[0.07, ..., 0.1]	[4]	[0.0, ..., 0.01]	[5]
[0.8, ..., 0.1]	[1]	[0.02, ..., 0.1]	[6]	[0.0, ..., 0.2]	[1]
...
[0.01, ..., 0.3]	[6]	[0.2, ..., 0.06]	[2]	[0.02, ..., 0.3]	[6]

Tableau 1 – Architecture de l'objet contenant les informations de la base de données

2^e expérimentation - CNN

La compression des données pour la 2^e expérimentation est beaucoup plus simple puisqu'il n'est pas nécessaire d'extraire un vecteur de caractéristiques pour chaque visage. Cette opération sera pris en charge par le réseau de neurones convolutif. En somme, il suffit de faire la lecture de l'image convertie en niveau de gris. Une fois toutes les images lues, l'algorithme utilise la même forme pour l'ensemble final de données qu'à la première expérimentation. La seule différence sera la longueur des vecteurs contenue à l'intérieur de la section "data" qui est de 100 x 100, soit 10 000 au lieu de 944.

2.1.2 – Entrainement et classification

1^{ère} expérimentation - LBP & SVM

Puisque toutes les données ainsi que leur étiquette correspondante sont désormais dans un fichier sérialisé, la première étape consiste à désérialiser le fichier et sauvegarder l'information dans deux vecteurs. Un premier vecteur d'information est utilisé pour entrainer le classifieur à vecteur de support et un deuxième est utilisé pour tester le même classifieur. La réalisation SVM utilisée lors de cette expérimentation est celle offerte par le module Python "scikit-learn". Pour trouver les hyperparamètres optimaux, une simple recherche par grille est utilisée. Le classifieur le plus performant a été réalisé avec les hyperparamètres suivants :

Noyaux : Fonction à base radial

C = 8.85

Gamma = 0.23

Noyaux : Linéaire

C = 1.43

2^e expérimentation - CNN

Tout comme la première expérimentation, la première étape de l'entraînement consiste à désérialiser le fichier où toutes les informations sont contenues et de sauvegarder ces informations dans deux vecteurs. Encore une fois, un premier vecteur d'information est utilisé pour entrainer le réseau de neurones et un deuxième est utilisé pour tester le réseau. Pour créer le réseau neuronal convolutif, l'implémentation fournie par la bibliothèque logicielle "Keras" a été utilisée. De plus, la version GPU de "TensorFlow" a aussi été utilisée comme support pour

“Keras”. Une fois la description du réseau terminée, il suffit de compiler le réseau et de l’entraîner.

Entrée (100x100x1)
Convolution 3x3 - 32
<i>Max pool</i>
Convolution 3x3 - 64
<i>Max pool</i>
Convolution 3x3 - 128
<i>Max pool</i>
Convolution 3x3 – 256
Convolution 3x3 - 256
FC-256
FC-256
FC-7

Tableau 2 – Architecture du CNN utilisée pour la 2^e expérimentation

2.2 - Description algorithmique

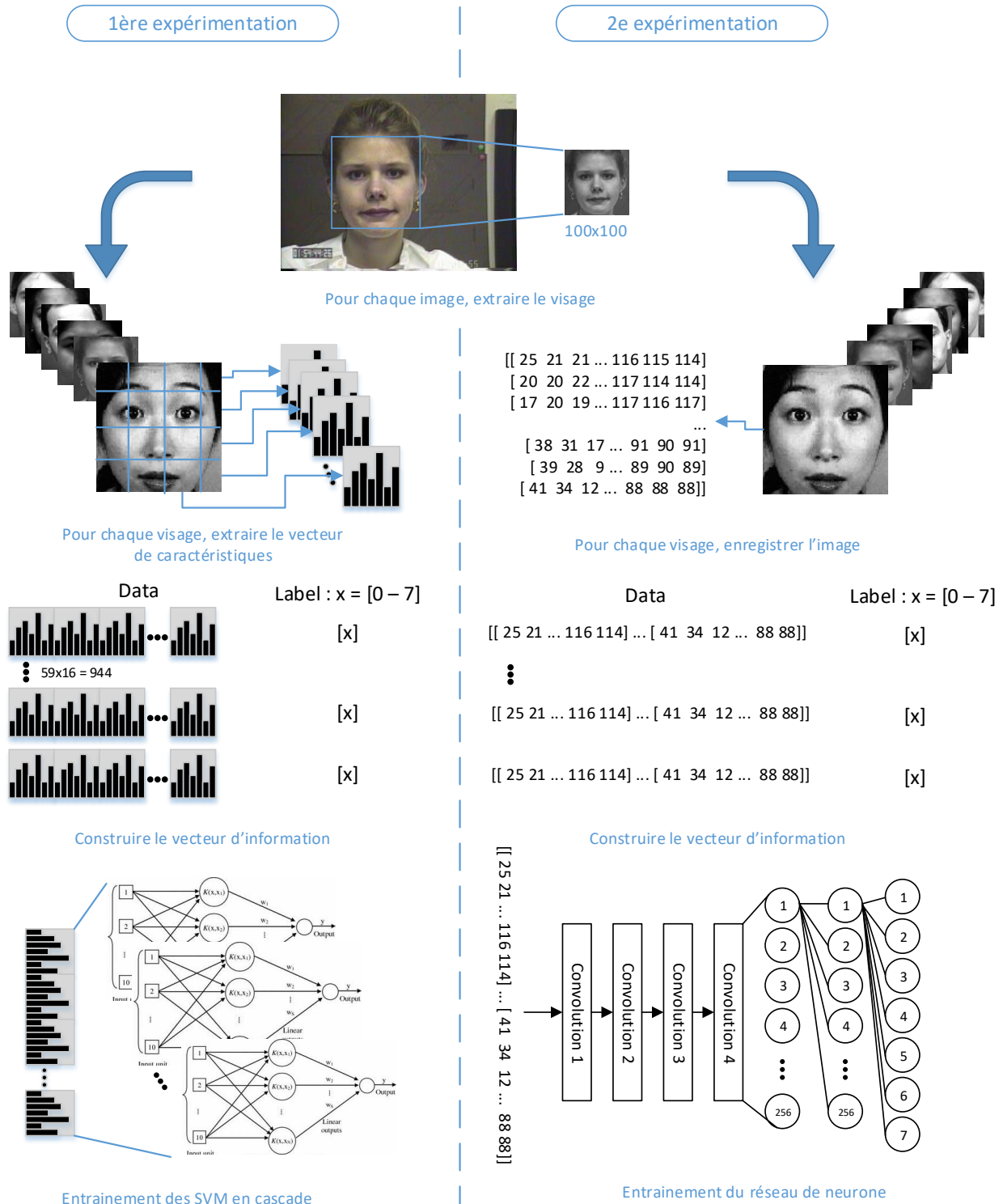


Figure 3 – Description algorithmique des deux systèmes

3 - Méthodologie expérimentale

3.1 - Base de données

Comme mentionné brièvement précédemment, la base de données utilisée pour les deux expérimentations est "Cohn-Kanade AU-Coded Expression Database". Il est important de souligner que la 2^e version de la base de données a été utilisée. Cette dernière contient 123 sujets différents qui totalisent 592 séquences d'expressions. Malheureusement, toutes les expressions contenues dans la base de données ne contiennent pas nécessairement d'étiquette. Ainsi, le nombre total d'expressions étiquetées revient à 326. Chaque expression contient en moyenne une séquence d'environ 12 images, la première étant un visage neutre et la dernière étant l'expression maximale. Deux méthodes bien distinguées ont été utilisées pour sélectionner les différentes images à traiter. Dans un premier temps, toutes les images étiquetées ont été traitées. Pour différencier entre une image neutre et une ayant une émotion, une limite arbitraire de 40% a été établie. La deuxième méthode consiste à considérer uniquement la première image de chaque séquence comme une image neutre et la dernière comme une image représentant l'émotion de l'étiquette.

<i>Émotion</i>	<i>Première méthode</i>	<i>Deuxième méthode</i>
Neutre	1292	327
Colère	776	45
Mépris	191	18
Dégoût	691	59
Peur	418	25
Joie	1034	69
Tristesse	422	27
Surprise	1052	83

Tableau 3 – Résumé du nombre d'images traitées selon la technique utilisée

Première méthode

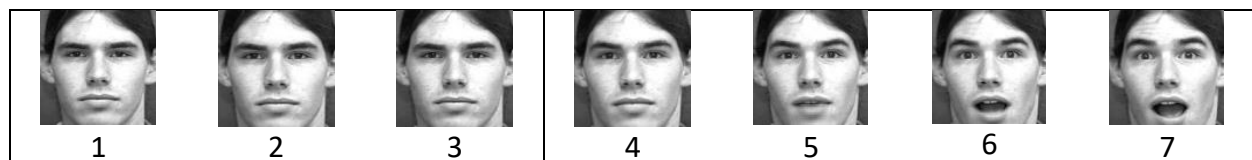


Figure 4 – Première méthode de tri de la base de données

Images considérées comme étant neutres : 1, 2, 3

Images considérées comme étant une émotion surprise: 4, 5, 6, 7

Deuxième méthode

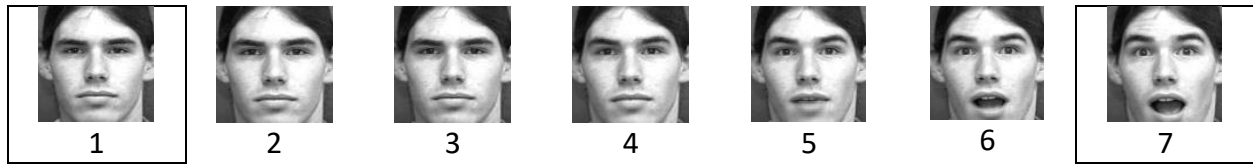


Figure 5 - Deuxième méthode de tri de la base de données

Images considérées comme étant neutre : 1

Images considérées étant une émotion surprise: 7

3.2 - Protocole d'évaluation

Pour assurer une comparaison juste entre les deux systèmes présentés précédemment, le protocole d'évaluation commence par l'utilisation de la même base de données tout au long des expérimentations. La base de données est séparée en deux portions de taille différente soit 90% et 10%. Le premier 90% de la base de données est utilisé pour entraîner les systèmes. Avec une telle approche, on s'assure qu'aucun système n'est présenté avec une base de données plus complexe. Pour le système basé sur un CNN, l'apprentissage est arrêté lorsque la précision sur la base de données de test est supérieure à 80%. Évidemment, pour le système à classifieur SVM, l'apprentissage est arrêté lorsque l'algorithme est complété.

Une fois les deux systèmes entraînés, il est possible de tester leur performance en présentant au système le 10% restant de la base de données initiale. En comparant les résultats prédits par chaque système à ceux présents sur les étiquettes, nous sommes en mesure de compiler toutes les mesures de performances présentées à la prochaine sous-section.

3.3 - Mesures de performances

Pour évaluer les performances des 2 systèmes présentés dans la section 2 du rapport, plusieurs mesures ont été utilisées. La première mesure de performance étudiée est la précision du système face à l'ensemble de la base de données de test. Cette mesure est essentielle et représente en somme le pourcentage de chance que le système a d'avoir la bonne prédiction émotionnelle lorsque présenté avec une image aléatoire. Une deuxième mesure de performance utilisée est la matrice de confusion. Cette matrice permet de visualiser la précision du système sur toutes les classes émotionnelles individuellement et non pas uniquement sur l'ensemble des classes. Finalement, une dernière mesure de performance sera utilisée soit le temps de traitement ainsi que le temps d'apprentissage des deux systèmes.

Uniquement applicable au 2^e système basé sur un CNN, une autre mesure de performance étudiée est la précision du système après chaque époque d'apprentissage. Cette métrique permet de visualiser la rapidité avec laquelle le système converge vers une solution acceptable.

4 - Résultats de simulation

4.1 - Présentation des résultats

	Architecture	Base de données	Émotions considérées	Augmentation des données	Noyau	Performance (Précision)
Article 1	LBP + SVM	CK+	Maximales	N/A	Linéaire	87%
	LBP + SVM	CK+	Maximales	N/A	RBF	88%
Article 2	BKStart + Softmax	FERC-2013	Maximales	Non	N/A	61%
	BKStart + Softmax	FERC-2013	Maximales	Oui	N/A	69%
Expérimentale	1 ^{er} système	CK+	Toutes	N/A	Linéaire	46%
	1 ^{er} système	CK+	Toutes	N/A	RBF	47%
	2 ^e système	CK+	Toutes	Non	N/A	56%
	2 ^e système	CK+	Toutes	Oui	N/A	60%
	1 ^{er} système	CK+	Maximales	N/A	Linéaire	81%
	1 ^{er} système	CK+	Maximales	N/A	RBF	83%
	2 ^e système	CK+	Maximales	Non	N/A	78%
	2 ^e système	CK+	Maximales	Oui	N/A	83%

Tableau 4 – Synthèse de la précision des expérimentations

Architecture	Étape du traitement	Temps de calcul (sec)
1 ^{er} système	Prétraitement de la base de données	289.7
	Compression de la base de données	48.65
	Entraînement et classification	0.29
	Total:	338.64
2 ^e système	Prétraitement de la base de données	289.7
	Compression de la base de données	5.32
	Entraînement et classification (Accélération GPU)	19.73
	Total:	314.75

Tableau 5 – Synthèse des temps de calcul des deux expérimentations

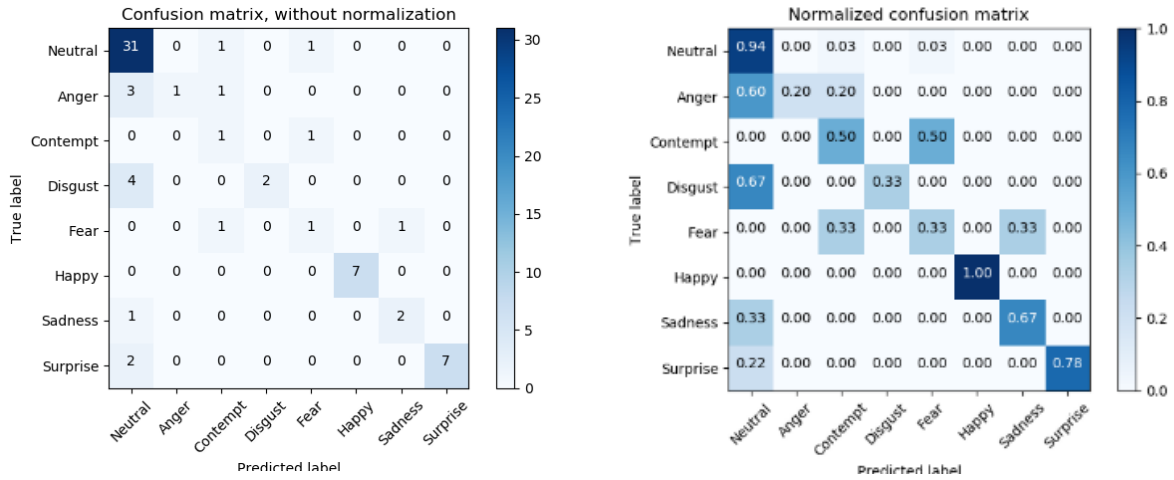


Figure 6 – Matrice de confusion de la première expérimentation

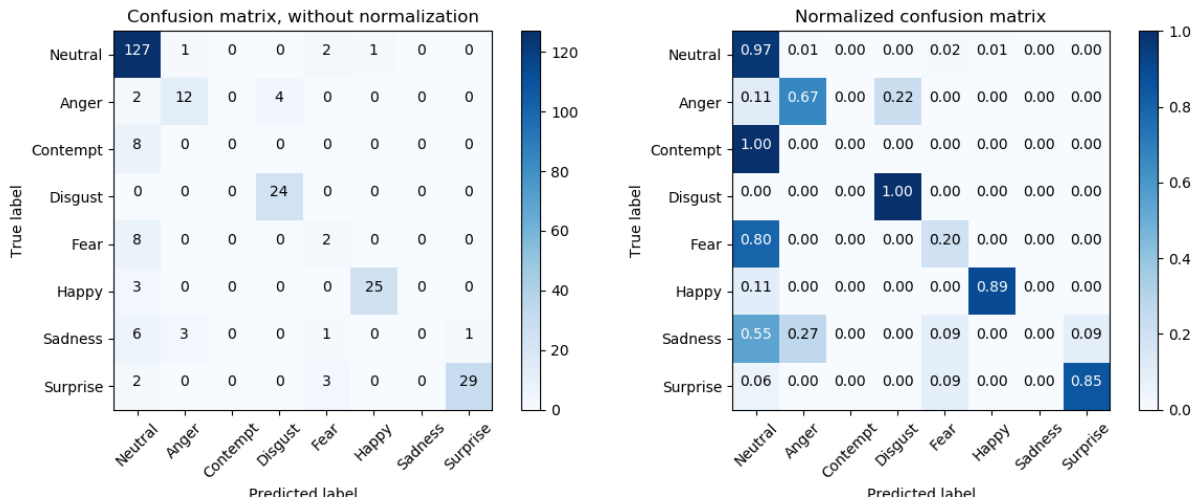


Figure 7 – Matrice de confusion de la deuxième expérimentation (données augmentées)

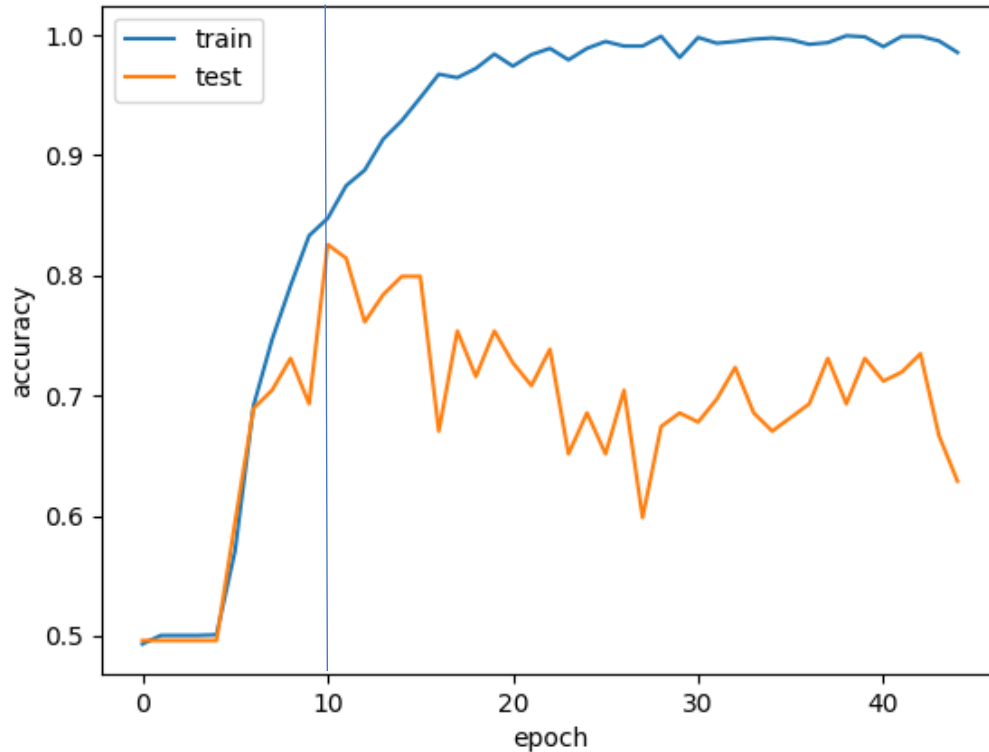


Figure 8 – Précision du deuxième modèle en fonction des époques

4.2 - Interprétation des résultats

Tel que décrit dans la section 3.1 du rapport, deux méthodes distinctes ont été utilisées pour sélectionner les images de la base données. La première méthode qui considère toutes les images de toutes les séquences émotionnelles produit des résultats peu désirables. Ces résultats sont néanmoins compréhensibles puisqu'il est extrêmement difficile pour les classifieurs de différencier un visage neutre d'un visage émotionnel. Les différences entre les visages de catégorie neutre et d'une autre catégorie sont tous simplement trop faibles.

D'une autre part, lorsque la deuxième méthode de sélection des données est employée, les résultats obtenus sont beaucoup plus semblables à ceux obtenues dans la littérature. Puisque les visages maximums seulement sont considérés, les différences entre ceux-ci sont beaucoup plus élevées et il est ainsi plus facile pour le classifieur de créer des bornes de décision. En observant le tableau 4, nous pouvons apercevoir que la précision maximale des deux systèmes est de 83% sur l'ensemble de test. En contrepartie, lorsque le deuxième modèle utilise des données non augmentées et que le premier modèle utilise un noyau linéaire, les performances diminuent légèrement soit, 81% et 78% respectivement. Le premier système, utilisant un classifieur SVM, n'obtient aucune augmentation significative de sa précision lorsqu'il utilise des données augmentées. Ainsi les résultats ne sont pas considérés.

Après une analyse brève des matrices de décision, on réalise que la plus grande difficulté du classifieur est de différencier un visage neutre à un visage émotionnelle. La grande majorité des erreurs restent la mauvaise classification d'un visage émotionnel comme étant un visage neutre. Il est toutefois intéressant de souligner que très peu d'erreurs surviennent dans la classification d'une émotion pour une autre émotion différente. De plus, le peu de données de test limitent notre habilité à bien analyser les performances des systèmes. Par exemple, pour le premier système, uniquement 3 images ont été présentées pour les émotions de peur et de tristesse.

À la figure 7, nous pouvons voir que les meilleures performances du deuxième système sont obtenues entre la 10^e et 12^e époque d'apprentissage dépendamment de la simulation. Lorsque le modèle dépasse 12 époques, un surapprentissage considérable est observable. Il est important de mentionner que ce surapprentissage survient malgré les deux couches "dropout" ajoutées dans le réseau de neurones.

5 - Conclusion

La question entreprise initialement dans la synthèse de littérature était de comparer un système plutôt classique à un modèle plus récent basé sur l'apprentissage profond. En utilisant la même base de données pour les deux modèles, il serait possible de comparer les performances de ceux-ci et d'en tirer une conclusion définitive. Après l'analyse des résultats, il est encore difficile de conclure avec certitude quelle approche est plus performante que l'autre. Il n'est pas possible de conclure que les deux systèmes sont de performance égale puisque nous ne pouvons regarder uniquement la précision du système comme barème d'évaluation. Malgré que les deux systèmes obtiennent une précision maximale de 83% sur l'ensemble de test, le temps et la puissance de calcul requis pour y parvenir varient grandement. Il est très important de mentionner que l'apprentissage et la classification du deuxième système a été effectuées à l'aide d'une carte graphique ce qui augmente considérablement la puissance de calcul. Sans l'utilisation d'une carte graphique, le temps d'entraînement était aux alentours d'une heure comparativement à 20 secondes comme présenté dans le tableau 5. Si la réalisation du premier modèle avait été parallélisée sur une carte graphique, le temps d'exécution de celui-ci aurait été bien inférieur au deuxième système.

Il est aussi important de souligner que la base de données utilisée favorise intrinsèquement le premier système. Le deuxième système basé sur un réseau neuronal convolutif dépend énormément de la quantité de données disponibles à l'apprentissage. Une petite base de données contenant 592 sujets limitent significativement le potentiel du classifieur utilisé dans le deuxième modèle. On ne peut cependant éliminer l'hypothèse qu'une plus grosse base de données pourrait augmenter la précision du deuxième système au point où elle pourrait dépasser celle du premier. Malheureusement, il n'existe pas présentement de large base de données étiquetée pour confirmer notre hypothèse.

Il est possible de tirer comme conclusion partielle qu'avec les bases de données présentement disponibles, les nouvelles techniques basées sur l'apprentissage profond n'offrent pas nécessairement de meilleures performances. Des performances similaires sont atteintes, mais avec une demande de puissance de calculs beaucoup plus élevée.

5.1 - Recommandations

Après la réalisation des deux expérimentations ainsi qu'une synthèse de littérature dans le domaine de la reconnaissance des émotions, il est évident que l'obstacle majeur à l'avancement du milieu est le manque de bases de données de taille et complexité appropriées. Les bases de données les plus populaires présentement sont très petites comparativement aux autres bases de données utilisées dans les autres sphères de l'apprentissage machine. Par exemple, la base de données de chiffre "EMNIST" contient 280,000 caractères comparativement à 592 sujets dans la base de données "Cohn-Kanade". Pour obtenir des

modèles dont les performances sont suffisantes pour des applications réelles, il sera nécessaire de posséder une base de données beaucoup plus complète que celles présentement disponibles.

Une autre option potentielle pour l'obtention de modèle plus performant est l'utilisation de méthodes mixtes. Par exemple, il est possible de combiner un vecteur de caractéristiques obtenu par LBP avec un deuxième vecteur obtenu par suivi de points géométriques. La combinaison de ces deux vecteurs permet d'extraire beaucoup plus d'informations du visage et augmente significativement les performances du système. Une étude³ obtient des résultats préliminaires de classification de 91.85% avec un tel système mixte. Plusieurs combinaisons de méthodes d'extraction de caractéristiques et de classification restent à découvrir. Avec des bases de données plus importantes et des modèles utilisant toutes les ressources disponibles, les performances des systèmes de reconnaissances émotionnelles seront significativement plus élevées dans le futur.

³ Datta, Samyak, et al. "Integrating Geometric and Textural Features for Facial Emotion Classification Using SVM Frameworks." *Advances in Intelligent Systems and Computing Proceedings of International Conference on Computer Vision and Image Processing*, 2016, pp. 619–628., doi:10.1007/978-981-10-2104-6_55.

Bibliographie

Logiciels

1. Mahakal, Krishna. "Facial Emotion Recognition." GitHub, github.com/mahakal/FacialEmotionRecognition.
2. Keras (Version 2.1.5) [Program documentation]. (n.d.). Retrieved from <https://keras.io/>
3. "Python Documentation." 3.6.5, [docs.python.org](https://docs.python.org/3.6.5/).
4. Python Modules: numpy, matplotlib, scikit-image, scikit-learn, opencv

Articles

5. Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(6), 1113-1133.
6. Tian, Y., Kanade, T., & Cohn, J. F. (2011). Facial expression recognition. In *Handbook of face recognition* (pp. 487-519). Springer London.
7. Whitehill, J., Bartlett, M. S., & Movellan, J. R. (2013). Automatic facial expression recognition. *Social Emotions in Nature and Artifact*, 88.
8. Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1), 39-58.
9. Ekenel, H. K., & Stiefelhagen, R. (2009, June). Why is facial occlusion a challenging problem? In *International Conference on Biometrics* (pp. 299-308). Springer, Berlin, Heidelberg.
10. Lien, J. J., Kanade, T., Cohn, J. F., & Li, C. C. (1998, April). Automated facial expression recognition based on FACS action units. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* (pp. 390-395). IEEE.
11. Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005, June). Recognizing facial expression: machine learning and application to spontaneous behavior. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 2, pp. 568-573). IEEE.
12. Zhang, Shiqing , et al. " Facial expression recognition based on local binary patterns and local fisher discriminant analysis." *WSEAS Transactions on Signal Processing*, vol. 8, no. 1, Jan. 2012, pp. 21–31.
13. Shan, Caifeng, et al. "Robust facial expression recognition using local binary patterns." *IEEE International Conference on Image Processing 2005*, 2005, doi:10.1109/icip.2005.1530069.
14. Sang, Dinh Viet, and Nguyen Van Dat. "Facial expression recognition using deep convolutional neural networks." *Knowledge and Systems Engineering (KSE)*, 2017 9th International Conference on. IEEE, 2017.
15. Fragopanagos, N., and John G. Taylor. "Emotion recognition in human–computer interaction." *Neural Networks* 18.4 (2005): 389-405.
16. Lopes, André Teixeira, et al. "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order." *Pattern Recognition* 61 (2017): 610-628.

17. El Kaliouby, Rana, and Peter Robinson. "Real-time inference of complex mental states from facial expressions and head gestures." Real-time vision for human-computer interaction. Springer, Boston, MA, 2005. 181-200.
18. Prince, E. B., Martin, K. B., & Messinger, D. S. Facial Action Coding System.
19. Datta, Samyak, et al. "Integrating Geometric and Textural Features for Facial Emotion Classification Using SVM Frameworks." Advances in Intelligent Systems and Computing Proceedings of International Conference on Computer Vision and Image Processing, 2016, pp. 619–628., doi:10.1007/978-981-10-2104-6_55.

Autres

20. Lynch, Vincent. "Synthèse De La Littérature, Reconnaissance et classification d'émotions." SYS843.