FORMATTING INSTRUCTIONS FOR APS360 PROJECT BASED ON ICLR CONFERENCE FORMAT

Hecan Zhou
Student# 1006069326
zhouheca@mail.utoronto.ca

Author Two
Student# 1005678901
author2@mail.utoronto.ca

Author Three
Student# 1005678901
author3@mail.utoronto.ca

Author Four
Student# 1005678901
author4@mail.utoronto.ca

ABSTRACT

This template should be used for all your project related reports in APS360 course.
– Write an abstract for your project here. Please review the **First Course Tutorial** for a quick start —-Total Pages: 8

1 INTRODUCTION

Facial expression detection has many different applications such as marketing and psychology, human-computer interaction. Leveraging deep learning, our project aims to develop a facial expression detection model capable of accurately identifying and categorizing emotions in static facial images and dynamic image sequences. For marketing purposes, facial expression detection can help personalize content delivery based on users' emotional responses. An example would be to continue showing content that has regularly made a user "happy". For psychology purposes, facial expression detection can help with mental health diagnosis such as detecting signs of depression, anxiety, trauma. Deep learning has demonstrated extraordinary success in image classification. In essence, this is an image classification problem by using facial features in order to classify an emotion. Due to this, deep learning is well suited for a facial expression detection project.

2 BACKGROUND/RELATED WORK

The GitHub page discusses using Convolutional Neural Networks (CNNs) and transfer learning for facial recognition, focusing on classifying faces by specific metrics. Inspired by this, our project aims to adapt this framework to detect emotions like smiles, anger, and sadness in facial expressions. We plan to use the same facial image dataset from the recognition project, applying a new labeling scheme to annotate emotions for training our CNN. Our project's architecture could follow the facial recognition project, with modifications for emotion detection. This involves adjusting the neural network to capture the emotional expressions effectively. In essence, for our project, we could repurpose this facial recognition technology for emotion detection, using a shared dataset and transfer learning to develop a model that classifies facial expressions into specific emotions. This streamlined approach aims for efficiency and accuracy in bridging facial recognition and emotion detection. We can take a pre-trained model that uses architecture such as ResNet and apply adjustment to make it fit for our usage. There also needs to be a pre-processing step that our team could develop and look into to prepare all data input for better recognition, base on paper from deep learning survey, a preprocessing step to align and highlight facial structure and area allows for the data input to have better quality and accuracy, leading to trained model that has better prediction and less losses Citation: https://github.com/jeffprosise/Deep-Learning/blob/master/Facial

James: The DISFA dataset comprises 27 videos, totaling 130,788 frames of facial expressions, and focuses on spontaneous facial expressions to more accurately capture the natural nuances of these

expressions. It serves as a dataset for models to train on. The author of this paper annotated the dataset based on the Facial Action Coding System (FACS), marking action units (AUs) for different levels of intensity. For our purposes, we can utilize these intensity annotations to develop new models that recognize not only the presence of an emotion but also its intensity. This approach enhances the classification problem by allowing the model to capture more intricate details of human emotions. https://paperswithcode.com/dataset/disfa The RAF-DB consists of 29,672 facial images tagged with basic or compound expressions, collected under real-world conditions. It encompasses a wide range of variability in age, gender, ethnicity, head poses, lighting conditions, and occlusions. While the DISFA dataset provides depth, the RAF-DB offers breadth. It is instrumental in training models to identify facial features in photos that represent real-world scenarios. By utilizing this dataset, we can explore how occlusions, lighting variations, and head poses affect the recognition process. This, in turn, increases the reliability of the model in real-world applications. https://paperswithcode.com/dataset/raf-db Alan: The following paper goes through many different models used in Facial Expression Recognition (FER). Some of which are described below.

Network Ensemble methods in (FER) involve the integration of spatial and temporal information from various networks and combining their outputs to improve performance. One example is a multi-channel network that extracted spatial and temporal information from emotion-expressing faces and changes between emotional and neutral faces.

Deep Spatiotemporal Networks are specifically designed to capture both spatial and temporal dependencies in consecutive frames, making them well-suited for FER tasks on dynamic sequences. These networks typically combine architectures such as Recurrent Neural Networks (RNNs) and Convolutional 3D (C3D) models. RNNs capture temporal dynamics of the learned features. C3D extends traditional CNNs by adding a 3D filter over very short video clips.

https://ieeexplore.ieee.org/abstract/document/9039580

3 Data Processing

4 Architecture

We plan on using a convolutional neural network with max pooling to reduce the spatial dimensions of the feature map. Padding should be used to preserve the edges of the images where distinct facial features may be located. For the activation function, ReLU will be used.

5 Baseline model

In the development of a Convolutional Neural Network (CNN) for face detection and facial recognition of facial expressions, our team plans to leverage established baseline models that utilize CNN architectures, such as AlexNet and ResNet. These will serve as a solid foundation for comparison, enabling us to test and benchmark our results effectively. Additionally, we will consider utilizing models pre-trained on architectures like VGG, time permitting. Employing pre-trained models based on ResNet or VGG as the backbone architecture allows us to use datasets such as the FERDataset or VGGFace2. This approach facilitates the evaluation of the impact of hyperparameter modifications, as well as the effects of increasing or decreasing the number of layers in our architecture, to fine-tune and adjust our model for enhanced accuracy.There are pre-existing models available in open-source GitHub repositories that have been trained with ResNet. For instance, the FaceNet project hosted at https://github.com/davidsandberg/facenet offers a pre-trained model using ResNet, boasting high accuracy and providing valuable data for comparison. Furthermore, the project available at https://github.com/WuJie1010/Facial-Expression-Recognition.Pytorch presents models trained with VGG19 and ResNet18, offering an additional benchmark for our results

6 Ethical consideration

The primary ethical concern for the topic of this project involves privacy and consent. Our model requires the input of photos or videos, which contain facial data that is inherently personal and sensitive. The collection and analysis of individuals' images should be conducted with consent. We

must ensure that individuals are fully informed about the data being collected, its intended use, and with whom it will be shared. This is particularly important in the current era, given the potential misuse of deepfake technology and other tools that can be exploited for impersonating individuals.

Another significant ethical issue is the potential for inherent bias within the detection models. These biases often come from imbalanced training datasets that fail to adequately represent minorities within the human population across dimensions such as race, gender, age, and culture. Quoting Joy Buolamwini's findings in "The Coded Gaze," it is evident that artificial intelligence systems, especially those involving facial analysis technologies, can reflect the biases of their creators and the datasets they are trained on. Many models have been shown to exhibit lower accuracy for women of color, highlighting a troubling discrepancy. This issue underscores the importance of fairness and equity in the field and points to the need for increased transparency and accountability within the machine learning community. To mitigate these issues, it is essential to utilize a diverse and representative dataset during the training phase. Ensuring that the dataset includes sufficient representation from various demographic groups and ethnicities is vital to achieving a fair training process. This approach guarantees that the resulting model performs consistently across all individuals without inherent biases.

7 PROJECT PLAN

8 LINKED TO GITHUB

https://github.com/hecan1234/Aps360Project

The format for the submissions is a variant of the ICLR 2022 format. Please read carefully the instructions below, and follow them faithfully. There is a **9 page** limit for the main text. References do not have any limitation. This is also ICLR's standard length for a paper submission. If your main text goes to page 10, a -20% penalty would be applied. If your main text goes to page 11, you will not receive any grade for your submission.

8.1 STYLE

Papers to be submitted to APS360 must be prepared according to the instructions presented here.

Authors are required to use the APS360 LaTeX style files obtainable at the APS360 website on Quercus. Tweaking the style is not permitted.

8.2 RETRIEVAL OF STYLE FILES

The file `APS360_Project.pdf` contains these instructions and illustrates the various formatting requirements your APS360 paper must satisfy. Submissions must be made using LaTeX and the style files `iclr2022_conference.sty` and `iclr2022_conference.bst` (to be used with LaTeX2e). The file `APS360_Project.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in sections 8, 9, and 10 below.

9 GENERAL FORMATTING INSTRUCTIONS

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, in small caps and left-aligned. All pages should start at 1 inch (6 picas) from the top of the page.

Authors' names are set in boldface, and each name is placed above its corresponding address. The lead author's name is to be listed first, and the co-authors' names are set to follow. Authors sharing the same address can be on the same line.

Please pay special attention to the instructions in section 10 regarding figures, tables, acknowledgments, and references.

There will be a strict upper limit of 9 pages for the main text of the initial submission, with unlimited additional pages for citations.

10 HEADINGS: FIRST LEVEL

First level headings are in small caps, flush left and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

10.1 HEADINGS: SECOND LEVEL

Second level headings are in small caps, flush left and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

10.1.1 HEADINGS: THIRD LEVEL

Third level headings are in small caps, flush left and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

11 CITATIONS, FIGURES, TABLES, REFERENCES

These instructions apply to everyone, regardless of the formatter being used.

11.1 CITATIONS WITHIN THE TEXT

Citations within the text should be based on the `natbib` package and include the authors' last names and year (with the "et al." construct for more than two authors). When the authors or the publication are included in the sentence, the citation should not be in parenthesis using `\citet{}` (as in "See **?** for more information."). Otherwise, the citation should be in parenthesis using `\citep{}` (as in "Deep learning shows promise to make progress towards AI (**?**).").

The corresponding references are to be listed in alphabetical order of authors, in the REFERENCES section. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

To cite a new paper, first, you need to add that paper's BibTeX information to `APS360_ref.bib` file and then you can use the `\citep{}` command to cite that in your main document.

11.2 FOOTNOTES

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).[2]

11.3 FIGURES

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

[1] Sample of the first footnote

[2] Sample of the second footnote

Table 1: Sample table title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

Figure 1: Sample figure caption. Image: ZDNet

11.4 TABLES

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

12 DEFAULT NOTATION

In an attempt to encourage standardized notation, we have included the notation file from the textbook, *Deep Learning* **?** available at `https://github.com/goodfeli/dlbook_notation/`. Use of this style is not required and can be disabled by commenting out `math_commands.tex`.

Numbers and Arrays

a	A scalar (integer or real)
$\boldsymbol{a}$	A vector
$\boldsymbol{A}$	A matrix
A	A tensor
$\boldsymbol{I}_n$	Identity matrix with n rows and n columns
$\boldsymbol{I}$	Identity matrix with dimensionality implied by context
$\boldsymbol{e}^{(i)}$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position i
$\text{diag}(\boldsymbol{a})$	A square, diagonal matrix with diagonal entries given by $\boldsymbol{a}$
a	A scalar random variable
$\mathbf{a}$	A vector-valued random variable
$\mathbf{A}$	A matrix-valued random variable

Sets and Graphs

$\mathbb{A}$	A set
$\mathbb{R}$	The set of real numbers
$\{0, 1\}$	The set containing 0 and 1
$\{0, 1, \dots, n\}$	The set of all integers between 0 and n
$[a, b]$	The real interval including a and b
$(a, b]$	The real interval excluding a but including b
$\mathbb{A} \backslash \mathbb{B}$	Set subtraction, i.e., the set containing the elements of $\mathbb{A}$ that are not in $\mathbb{B}$
$\mathcal{G}$	A graph
$Pa_{\mathcal{G}}(\mathrm{x}_i)$	The parents of x_i in $\mathcal{G}$

Indexing

a_i	Element i of vector $\boldsymbol{a}$, with indexing starting at 1
a_{-i}	All elements of vector $\boldsymbol{a}$ except for element i
$A_{i,j}$	Element i, j of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{i,:}$	Row i of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{:,i}$	Column i of matrix $\boldsymbol{A}$
$\mathsf{A}_{i,j,k}$	Element (i, j, k) of a 3-D tensor A
$\mathsf{A}_{:,:,i}$	2-D slice of a 3-D tensor
a_i	Element i of the random vector $\mathbf{a}$

Calculus

$\frac{dy}{dx}$	Derivative of y with respect to x
$\frac{\partial y}{\partial x}$	Partial derivative of y with respect to x
$\nabla_{\boldsymbol{x}} y$	Gradient of y with respect to $\boldsymbol{x}$
$\nabla_{\boldsymbol{X}} y$	Matrix derivatives of y with respect to $\boldsymbol{X}$
$\nabla_{\mathbf{X}} y$	Tensor containing derivatives of y with respect to $\mathbf{X}$
$\frac{\partial f}{\partial \boldsymbol{x}}$	Jacobian matrix $\boldsymbol{J} \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \to \mathbb{R}^m$
$\nabla^2_{\boldsymbol{x}} f(\boldsymbol{x})$ or $\boldsymbol{H}(f)(\boldsymbol{x})$	The Hessian matrix of f at input point $\boldsymbol{x}$
$\int f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral over the entire domain of $\boldsymbol{x}$
$\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral with respect to $\boldsymbol{x}$ over the set $\mathbb{S}$

Probability and Information Theory

$P(\mathrm{a})$	A probability distribution over a discrete variable
$p(\mathrm{a})$	A probability distribution over a continuous variable, or over a variable whose type has not been specified
$\mathrm{a} \sim P$	Random variable a has distribution P
$\mathbb{E}_{\mathrm{x} \sim P}[f(x)]$ or $\mathbb{E} f(x)$	Expectation of $f(x)$ with respect to $P(\mathrm{x})$
$\mathrm{Var}(f(x))$	Variance of $f(x)$ under $P(\mathrm{x})$
$\mathrm{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(\mathrm{x})$
$H(\mathrm{x})$	Shannon entropy of the random variable x
$D_{\mathrm{KL}}(P \Vert Q)$	Kullback-Leibler divergence of P and Q
$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution over $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

Functions

$f : \mathbb{A} \to \mathbb{B}$	The function f with domain $\mathbb{A}$ and range $\mathbb{B}$
$f \circ g$	Composition of the functions f and g
$f(\boldsymbol{x}; \boldsymbol{\theta})$	A function of $\boldsymbol{x}$ parametrized by $\boldsymbol{\theta}$. (Sometimes we write $f(\boldsymbol{x})$ and omit the argument $\boldsymbol{\theta}$ to lighten notation)
$\log x$	Natural logarithm of x
$\sigma(x)$	Logistic sigmoid, $\frac{1}{1 + \exp(-x)}$
$\zeta(x)$	Softplus, $\log(1 + \exp(x))$
$\Vert \boldsymbol{x} \Vert_p$	L^p norm of $\boldsymbol{x}$
$\Vert \boldsymbol{x} \Vert$	L^2 norm of $\boldsymbol{x}$
x^+	Positive part of x, i.e., $\max(0, x)$
$\mathbf{1}_{\mathrm{condition}}$	is 1 if the condition is true, 0 otherwise

13 FINAL INSTRUCTIONS

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the REFERENCES section; see below). Please note that pages should be numbered.

AUTHOR CONTRIBUTIONS

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

ACKNOWLEDGMENTS

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.