Facial Action Units Detection using Transfer Learning

\*Note: Sub-titles are not captured in Xplore and should not be used

line 1: 1st Given Name Surname   
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 4th Given Name Surname  
line 2: *dept. name of organization*  
*(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: 2nd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 5th Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: 3rd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 6th Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

*Abstract*—นำเสนอ Facial Action Units detection model โดยใช้ transfer learning ร่วมกับการทำ feature engineering ด้วยการสกัด 3 feature ได้แก่ 1. Histogram of Oriented Gradient (HOG) 2. Face landmark 3. Angle of center key point เพื่อเพิ่มความแม่นยำในการตรวจจับ โดยทำการเปรียบเทียบโมเดลที่นำเสนอกับโมเดลอื่นๆ ทั้งด้าน acc และ time

Keywords— Facial Action Units, Transfer learning, facial keypoint, face landmark, Emotion

# Introduction

การประเมินอารมณ์ความรู้สึกของผู้ที่ต้องการพบนักจิตวิทยา เริ่มแรกมีความพยายามเข้าใจถึงกลไกที่สมองสามารถรับรู้อารมณ์จากสีหน้า โดยมีสมมติฐานว่าน่าจะเกิดจากการเรียนรู้จดจำการเคลื่อนไหวของใบหน้า ในปี 1978 นักจิตวิทยาชาวอเมริกัน Ekman และ Friesen ได้จำแนกการเคลื่อนไหวบนใบหน้าออกเป็น 46 หน่วยกระทำ (action unit : au) [1] ซึ่งเรียกระบบนี้ว่า Facial Action Coding System (FACS) เป็นระบบที่ใช้วิธีใส่หมายเลขกำกับการเคลื่อนไหวกล้ามเนื้อแต่ละส่วน 46 ส่วนบนใบหน้า ระบบนี้ได้รับการยอมรับให้เป็นมาตรฐานในการศึกษาการแสดงออกใบหน้าตั้งแต่นั้นมา [2]

ในช่วงแรกการใช้ FACS สำหรับการศึกษาวิจัยทำได้อย่างจำกัดเนื่องจากผู้ประเมินต้องได้รับการฝึกอบรมจนได้ประกาศนียบัตร ผลที่ประเมินจึงจะน่าเชื่อถือ จนกระทั้งในปี 2011 Hamm J ได้พัฒนาอัลกอริทึมเพื่อประเมิน FACS ขึ้น [3] หลังจากนั้นจึงมีการศึกษาวิจัย FACS ในทาง behavioral medicine อย่างกว้างขวางมากขึ้น โดยในช่วงแรกๆของ FACS ถูกนำมาใช้ในการแยกแยะอารมณ์ เช่น angry, happy, disgust เป็นต้น ยกตัวอย่าง การยิ้ม 2 แบบ แบบแรกเป็นการยิ้มแบบพนักงานตอนรับบนเครื่องบิน มุมปากยกขึ้น ใช้กล้ามเนื้อส่วนที่เรียกว่า กล้ามเนื้อไซโกมาติคัสเมเจอร์ แต่ใบหน้าส่วนอื่นไม่ขยับ รอยยิ้มจึงดูเสแสร้ง เป็นรอยยิ้มซึ่งไม่มีสีหน้าใดๆ ประกอบอยู่ในนั้น ตามระบบ FACS รอยยิ้มซึ่งใช้กล้ามเนื้อส่วนนี้มีรหัสเป็น AU12 ในขณะที่การยิ้มแบบจริงใจตาม FACS คือ AU12 บวก AU6 หมายถึงใบหน้าขยับเคลื่อนไหว รวมไปถึงด้านนอกของกล้ามเนื้อออร์บิคูลาริส โอคูลี แก้มยกสูงจนเห็นตีนการอบๆ ดวงตาชัดเจน

ซึ่งงานวิจัยส่วนใหญ่ในปัจจุบันมีการนำ Action unit ไปใช้ในการตีความหมายของอารมณ์

1. Emotion description in terms of action units.

| Emotion | Criteria |
| --- | --- |
| Angry | AU23 + AU24 |
| Disgust | AU9 or AU10 |
| Fear | AU1+2+4, unless AU5 is of  intensity E then AU4 can be absent |
| Happy | AU12 |
| Sadness | Either AU1+4+15 or 11 must be present.  An exception is AU6+15 |
| Surprise | Either AU1+2 or 5 must be present and  the intensity of AU5 must not be stronger  than B |
| Contempt | AU14 must be present  (either unilateral or bilateral) |

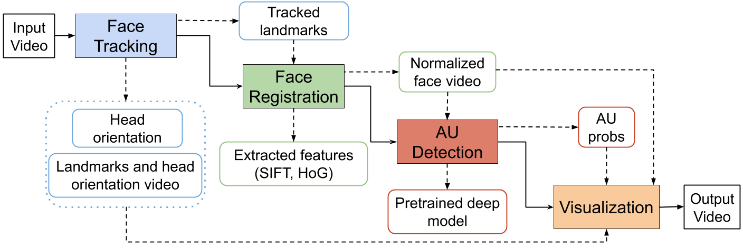
# Related Work

หลังจากได้ทำการศึกษาค้นคว้าในหลายๆงานวิจัยพบว่างานวิจัย [5] ได้สร้างแบบจำลองสำหรับการตรวจจับหน่วยกระทำ โดยทำการฝึกสอนแบบจำลองด้วยฐานข้อมูล CK+ โดยมีการใช้ Local Binary Pattern (LBP) เพื่อสกัด features และ textures ที่สำคัญของรูป ทำการลด dimensionality ของข้อมูลโดยการใช้ PCA / LDA / Adaboost / Gentleboost ฝึกสอนโมเดล SVM classifier ด้วย RBF/Linear kernel และมีการใช้เทคนิค optimization เพื่อหาค่า parameter ที่ดีที่สุดของสมการ

# FACIAL ACTION UNITS DeTEction

## ภาพรวมการกระบวนการ action unit detection

การตรวจจับหน่วยกระทำมีอยู่ด้วยกัน x ขั้นตอน ได้แก่ 1. การตรวจจับส่วนที่เป็นใบหน้า 2.นำภาพส่วนของใบหน้าที่ตรวจจับได้ไปหา facial landmark/key point จำนวน 68 จุด ดังแสดงในรูปด้านล่าง



รูปตัวอย่าง

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

## Face detection

การตรวจจับภาพใบหน้าปัจจุบันมีหลากหลายวิธีการซึ่งมีข้อดีข้อเสียแตกต่างกันไปทั้งในแง่ของความแม่นยำและความเร็วในการประมวลผล ยกตัวอย่างเช่น วิธีการแรก เป็นการแยกใบหน้าออกจาก background มี feature input เป็นการเปลี่ยนแปลงไล่ระดับของสีในภาพ (Histogram of Oriented Gradient: HOG) และโมเดลที่ใช้ในการ classify ส่วนใบหน้า และ background คือ Support Vector Machine (SVM) ซึ่งวิธีการนี้มีข้อเสียตรงที่การถ่ายภาพในที่มีแสงน้อยอาจมีผลกระทบต่อความแม่นยำได้ วิธีการต่อมาคือ … เป็นต้น

ปัจจุบันมี libraries สำหรับการ face detection ในหลายๆโมเดลซึ่งในแต่ละ library จะมีความแตกต่างกันในสองปัจจัยหลักก็คือ acc และ time ดังแสดงในตารางที่ x ด้านล่าง เครื่องหมายถูกหมายถึงมีจุดเด่นในด้านนั้น

1. Face detection

| No | Face detection | | |
| --- | --- | --- | --- |
| Libraries | Time | acc |
| 1 | Dlib | ✓ |  |
| 2. | MTCNN |  |  |
| 3. | FaceBoxes |  |  |
| 4. | RetinaFace | ✓ | ✓ |
| 5. | Haar cascade |  |  |

## Facial landmark

การประเมินการขยับของกล้ามเนื้อใบหน้าหรือการอธิบายการเคลื่อนที่ส่วนต่างๆของใบหน้าต้องอาศัยการกำหนดจุดอ้างอิง (landmark) และการกำหนด action unit ต้องอาศัยจุดอ้างอิงนั้น ดังนั้น facial landmark จึงเป็นหนึ่งใน feature ที่สำคัญสำหรับการตรวจจับหน่วยกระทำ ปัจจุบันมี libraries สำหรับการ detect face landmark ในหลายๆโมเดลซึ่งในแต่ละ library จะมีความแตกต่างกันในสามปัจจัยหลักก็คือ acc , size และ time ดังแสดงในตารางที่ x ด้านล่าง เครื่องหมายถูกหมายถึงมีจุดเด่นในด้านนั้น

1. Face Landmark

| No | Face Landmark | | | | |
| --- | --- | --- | --- | --- | --- |
| Libraries | Time | acc | size |
| 1 | Dlib | ✓ |  | ✓ |
| 2 | Openpose |  | ✓ |  |
| 3 | Mediapipe | ✓ | ✓ | ✓ |
| 4 | Pre-Trained model  (Xception, ResNet,  VGG) |  | ✓ |  |
| 5 | stasm | ✓ |  | ✓ |
| 6 | PFLD |  |  |  |
| 7 | MobileFaceNet |  |  | ✓ |
| 8 | MobileNet |  |  |  |

สำหรับตัว Open pose เราทดสอบการทดสอบทั้งที่เป็น State of the art, Caffe, TensorFlow และ Chainer model

## Action Unit Detection (classification)

การตรวจจับ action unit สามารถตรวจจับได้จากการเปลี่ยนของ face landmarks ต่างๆบนใบหน้า ปัจจุบันมี libraries สำหรับการตรวจจับ action unit อยู่ไม่กี่โมเดล ดังแสดงในตารางที่ x ซึ่งปัจจุบันยังมีความแม่นยำที่ไม่สูงมาก จากการศึกษาพบว่า หลายๆงานวิจัยกำลังปรับปรุงความแม่นยำของการตรวจจับ action unit

วิธีการที่นำเสนอได้สร้าง feature ขึ้นมาใหม่สอง feature ได้แก่ ระยะห่างระหว่างแต่ละ face landmark กับ center point ของทุก landmark และ มุมของแต่ละ face landmark กับ center ของทุก face landmark ซึ่ง feature ทั้งหมดที่นำเสนอจะอยู่ในรูป <x, y, distance from center, angle relative to center> ซึ่ง feature เหล่านี้จะช่วยลดความผิดผลาดที่เกิดจากการเอนเอียงของใบหน้า

1. Action unit detection

| No | Action unit detection | | | | |
| --- | --- | --- | --- | --- | --- |
| Libraries | Time | acc | size |
| 1 | random forest |  |  |  |
| 2 | svm |  |  |  |
| 3 | logistic |  |  |  |
| 4 | JAANET |  |  |  |
| 5 | DRML |  |  |  |
| 6 | Perposed model |  |  |  |

## Regression model

Regression เป็นการบอกว่า frame นั้นมีความเปลี่ยนแปลงไปจาก frame ก่อนในระดับมากน้อยเพียงใด

# อธิบายข้อมูล

Dataset ที่ใช้ในงานวิจัยนี้คือ CK+ dataset เนื่องมี labelled AUs และ labelled basic emotions. Dataset นี้ประกอบไปด้วย 593 image sequences ของ 123 subjects. อนุกรมของรูปภาพทั้งหมดเริ่มที่การแสดงออกทางอารมณ์แบบปกติและสิ้นสุดด้วยการแสดงออกทางอารมณ์แบบ peak โดยการ label AU นั้นมีการ label 2 รูปแบบ ได้แก่ 1. Label ที่บอกว่า AU นั้นเกิดขึ้นหรือไม่ (classification) 2. Label ที่บอกความรุ่นแรงของ AU ที่เกิดขึ้น (regression)

มี 327 อนุกรมรูปภาพที่ถูก label สำหรับ basic emotions.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even.

## Units

* Use either SI

## Equations

The equations are an exception to the prescribed specifications of

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other

An excellent style manual for science writers is [7].

# การทดลอง

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Face detection

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference article

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Facial Landmark detection

ผลการทดลองจะเป็นการเปรียบเทียบ acc และ time แล้วเลือก face landmark model ที่เหมาะสมที่สุด

## Action Units Detection

ผลการทดสอบการตรวจจับ action unit บน CK+ dataset แสดงดังตารางด้านล่าง

#### Positioning Figures and Tables: Place figures and tables at the top

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis

# ผลลัพธ์และสรุป

แสดง heatmap layer สุดท้ายของ model

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. https://imotions.com/blog/facial-action-coding-system/
2. Facial Action Coding System [Internet]. Paul Ekman Group. [cited 2020 Jun 28]. Available from: <https://www.paulekman.com/facial-action-coding-system/>
3. Hamm J, Kohler CG, Gur RC, Verma R. Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. Journal of Neuroscience Methods 2011;200:237-56.
4. Kunz M, Seuss D, Hassan T, Garbas JU, Siebers M, Schmid U, et al. Problems of video-based pain detection in patients with dementia: a road map to an interdisciplinary solution. BMC Geriatrics [Internet]. 2017 Dec [cited 2018 Mar 22];17(1). Available from: http://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-017-0427-2
5. Y. Zhang, W. Dong, B. Hu and Q. Ji, "Classifier Learning with Prior Probabilities for Facial Action Unit Recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5108-5116, doi: 10.1109/CVPR.2018.00536.
6. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
7. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
8. K. Elissa, “Title of paper if known,” unpublished.
9. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
10. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
11. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.