

Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning

渋谷

目次

- ・感情認識業界の現状 1: マイクロソフトEmotion API
- ・感情認識業界の現状 2: 日銀総帥の会見の表情分析
- ・感情認識コンテスト EmotiW
- ・EmotiW2015に使用されたデータ
- ・EmotiW2015優秀チームの成果
- ・感情認識タスクの抱える問題点

マイクロソフト Emotion API



写真をアップロードすると自動的に顔を探して感情ラベルをつけてくれる。

Pythonなどのプログラムで大量の画像を認識させることも可能。

性能

Sighthound says that it has the world's "most accurate age detector, gender recognizer, and emotion classifier" It says it provides "significantly more accurate results than the Microsoft Cognitive Services API."

Its "**Real Age estimation**" **absolute error** is 5.76 against Microsoft Cognitive services API (7.62), which is 1.86 years more accurate. Its "**Emotion Recognition accuracy**" is 76.1 percent compared to Microsoft's 61.3 percent.

Its **gender recognition** is accurate 91 percent of the time compared to Microsoft Cognitive Services accuracy, which is accurate 90.86 percent of the time.

フィルター

- Emotion API
- Documentation
- Overview
- How to
- Quickstarts
- cURL
- C#
- Java
- JavaScript
- PHP
- Python**
- Ruby
- Tutorials
- Reference
- Resources

このコンテンツはお使いの言語では利用できません。英語版はこちらです。

Emotion API Python Quick Start

前 2017/05/23 読み取りに 2 分 • 共同作成者

Important

Video API Preview will end on October 30th, 2017. Try the new Video Indexer API Preview to easily extract insights from videos and to enhance content discovery experiences, such as search results, by detecting spoken words, faces, characters, and emotions. Learn more.

This article provides information and code samples to help you quickly get started using the Emotion API Recognize method with Python to recognize the emotions expressed by one or more people in an image.

Prerequisite

- Get your free Subscription Key [here](#)

Recognize Emotions Python Example Request

Copy the appropriate section for your version of Python and save it to a file such as test.py. Replace the "Ocp-Apim-Subscription-Key" value with your valid subscription key, add a URL to a photograph of a celebrity to the body variable, and change the REST URL to use the region where you obtained your subscription keys.

```
Python
##### Python 2.7 #####
import urllib, urllib2, base64

headers = {
    'Content-Type': 'application/json',
    'Ocp-Apim-Subscription-Key': '13hc7781f7e4b19b5fcd72a8df7156',
}

params = urllib.urlencode({
})

# Replace the example URL below with the URL of the image you want to analyze.
body = { 'url': 'http://example.com/picture.jpg' }

try:
    # NOTE: You must use the same region in your REST call as you used to obtain
    # your subscription key. For example, if you obtained your subscription keys from westcentralus,
    # use the URL below with westcentralus.
    conn = urllib2.HTTPSConnection('westus.api.cognitive.microsoft.com')
    conn.request('POST', '/emotion/v1.0/recognize?%s' % params, body, headers)
    response = conn.getresponse()
    data = response.read()
    print(data)
    conn.close()
except Exception as e:
    print("[Errno %s] %s" % (e.errno, e.strerror))

##### Python 3.2 #####
import http.client, urllib.parse, urllib.error, base64, sys

headers = {
    'Content-Type': 'application/json',
    'Ocp-Apim-Subscription-Key': '13hc7781f7e4b19b5fcd72a8df7156',
}

params = urllib.parse.urlencode({
})

# Replace the example URL below with the URL of the image you want to analyze.
body = { 'url': 'http://example.com/picture.jpg' }

try:
    # NOTE: You must use the same region in your REST call as you used to obtain
    # your subscription key. For example, if you obtained your subscription keys from westcentralus,
    # use the URL below with westcentralus.
    conn = http.client.HTTPSConnection('westus.api.cognitive.microsoft.com')
    conn.request('POST', '/emotion/v1.0/recognize?%s' % params, body, headers)
    response = conn.getresponse()
    data = response.read()
    print(data)
    conn.close()
except Exception as e:
    print(e.args)
#####
```

Recognize Emotions Sample Response

A successful call returns an array of face entries and their associated emotion scores, ranked by face rectangle size in descending order. An empty response indicates that no faces were detected. An emotion entry contains the following fields:

編集

共有

テーマ

目次

この記事の内容

Prerequisite

Recognize Emotions

Python Example

Request

Recognize Emotions

Sample Response

Pythonなどのサンプルプログラムも公式ページに載せている。

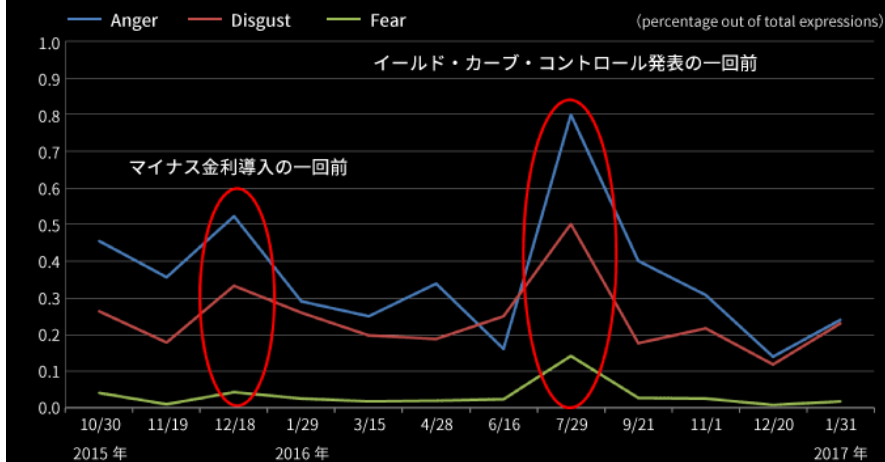
サンプルコードを改良して様々な条件で感情分析を行っているネット記事も多数。

日銀 黒田総帥の表情分析

政策変更前は「怒り」と「嫌悪」が上昇、変更後は「悲しみ」が低下
水門善之氏(野村証券金融経済研究所) 勇大地氏(米マイクロソフト)

AI が解析した黒田総裁の感情スコア

政策変更は「怒り」と「嫌悪」が上昇



出典: Yoshiyuki Suimon, Daichi Isami

REUTERS

AI が解析した黒田総裁の感情

政策変更後は「悲しみ」スコアが低下



出典: Yoshiyuki Suimon, Daichi Isami

REUTERS

まず、日経チャンネル[2]にて公表されている会見動画を約 0.5 秒ごとにスクリーンショットを撮り、解析の対象とする画像データを作成した。

本研究では Microsoft の Cognitive Service における表情認識アルゴリズムを感情値の計測に用いた[4]。そ

$$\text{喜びの割合} = \frac{\sum_{t=\text{Start}}^{\text{End}} (\text{喜び}_t)}{\sum_{t=\text{Start}}^{\text{End}} (\text{喜び}_t + \text{悲しみ}_t + \text{中立}_t + \dots)}$$

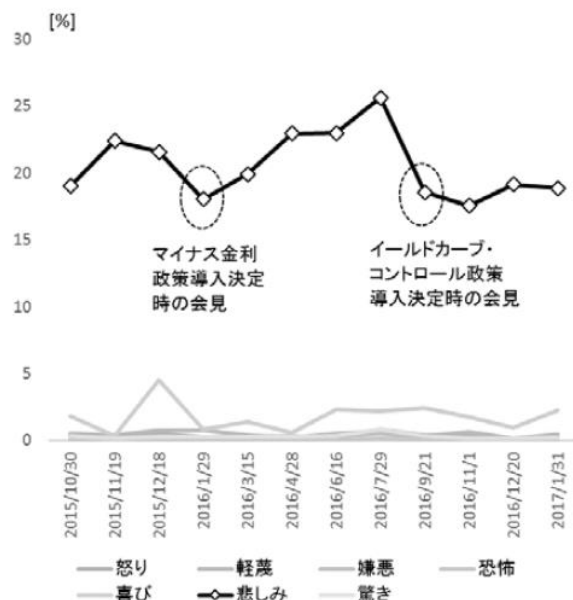


図 4: 政策変更後に落ち着きを見せる感情値

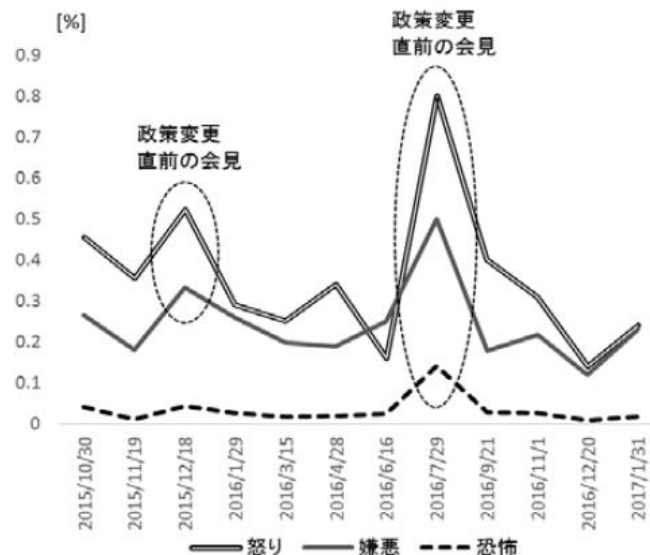


図 3: 政策変更前に見られる感情値の変化

例えば、前述したマイナス金利政策の導入や、イールドカーブ・コントロール政策の導入といった、それまでの金融政策のフレームワークを変更するような、重大な金融政策変更を行う直前の記者会見では、「怒り」や「嫌悪」の割合が高くなる傾向が見られた (図 3)。また、それらの金融政策変更に向けては、「悲しみ」の割合が上昇を続け、金融政策変更の決定を行った後の記者会見では、「悲しみ」の割合が低下する傾向も見られた (図 4)。

このことは、金融政策の変更前に既存の金融政策に対する問題意識の高まりがネガティブな表情として表れていた可能性と、また、金融政策変更によって、それが緩和されたことによる安堵が、表情に表れていた可能性を示していると言えよう。

もちろん、記者会見における総裁の表情の変化は、その発言の内容によるところも大きい点には注意が必要だ。ただし、通常、日銀総裁は公の場において、「怒り」や「嫌悪」といったネガティブな感情を露わにする場面は少ない。本研究では、人工知能技術を用いることで、表情の機微な変化を計量的に測定したことで、総裁の表情に表れたネガティブな感情の高まりを時系列で捉えることができたと考える。

実際、既存の金融政策に対する問題意識が高まることで、新たな金融政策への変更が施される点を踏まえると、政策変更前に、ネガティブな感情が高まることは、理にかなっている。

「悲しみ」と「中立」で98%以上の中で、怒りが0.1%から0.8%なったことをシグナルと呼べるのか・・・

留意事項

本稿は、著者の個人見解を表すものであり、野村證券株式会社および Microsoft Corporation の公式見解を表すものではありません。

EmotiW

(Emotion Recognition in the Wild Challenge)

一般的なシチュエーションでの感情認識のコンペティション

年によってルール(課題や使用するデータセット)が異なる。

2015は2種目

1. Audio-Video based emotion recognition challenge based on AFEW database [1]
2. Image based static facial expression recognition challenge based on SFEW database [2]

今回の論文は2の静止画の感情認識

Acted Facial Expressions in the Wild (AFEW)

聴覚障害者用字幕(SDH)とクローズドキャプション(CC)に基づいて収集された。

クリップに関する情報は、拡張可能なXMLスキーマに格納され、クリップのサブジェクトには、名前、俳優の年齢、キャラクターの年齢、性別、性別、人の表現、全体的なクリップエクспレッションなどの属性が付いている。

Static Facial Expressions in the Wild (SFEW)

AFEW配列からフレームを抽出し、配列のラベルに基づいて標識した。怒り、嫌悪感、恐怖、幸せ、悲しみ、驚き、中立の7つの基本的な表現にラベル付けされている。(Azureは8分類。喜び、中立、怒り、驚き、嫌悪感、軽蔑、悲しみ、恐怖)

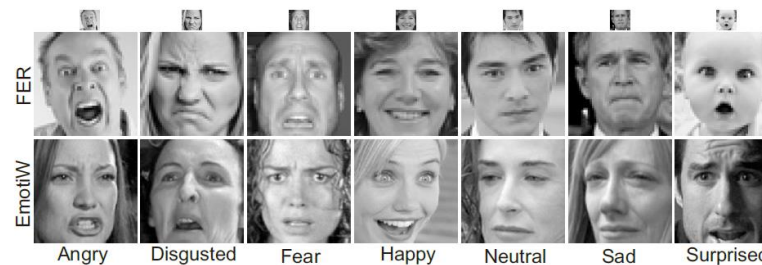


Figure 2: Comparison between the FER-2013 and EmotiW datasets. Top row: original size of the FER-2013 dataset (48×48 pixels). Middle row: upsampled FER-2013 dataset to 256×256 pixels. Bottom row: EmotiW dataset (256×256 pixels).

the eye-to-eye distance ($e2e$):
 The side boundaries were $0.62 \times e2e$
 (counted from the respective eye corner)
 the upper and lower boundaries were $0.9 \times e2e$ and $1.35 \times e2e$,



Organizers' cropping Our cropping

Figure 1: Comparison between our cropping method and the one provided by the organizers.

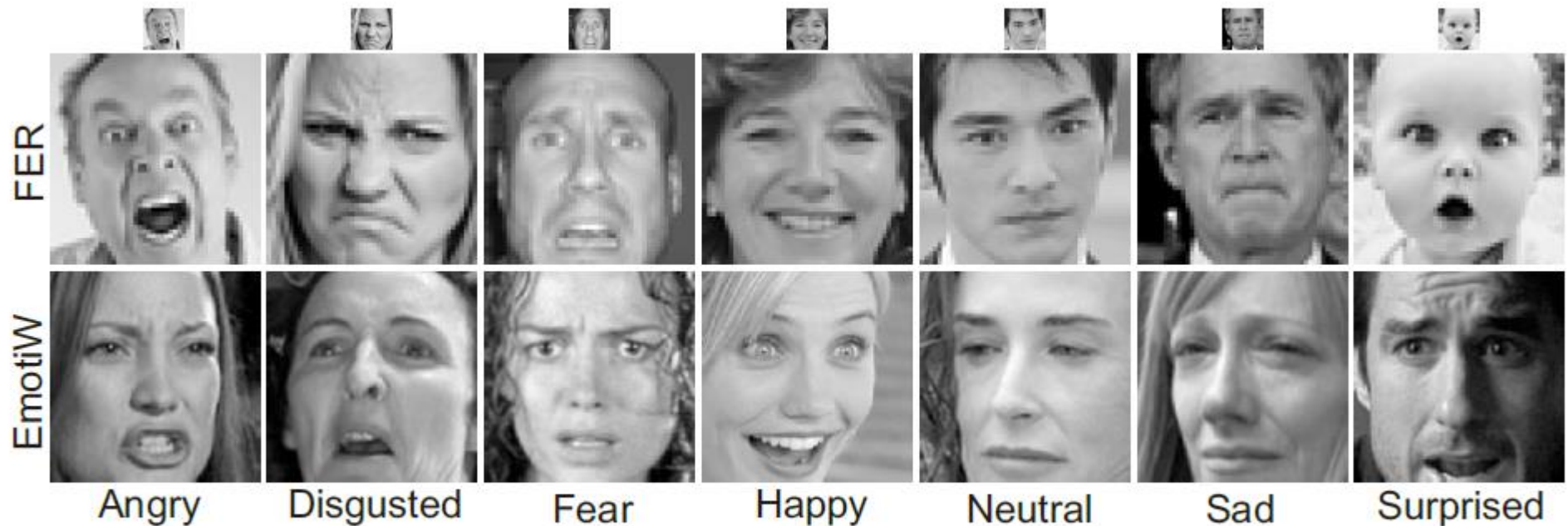


Figure 2: Comparison between the FER-2013 and EmotiW datasets. Top row: original size of the FER-2013 dataset (48×48 pixels). Middle row: upsampled FER-2013 dataset to 256×256 pixels. Bottom row: EmotiW dataset (256×256 pixels).

ネットワークのトレーニングには、過去のコンテストに使用された比較的大きなデータセットと、今回のコンテストの小さなデータセットを併用。(転移学習)
 転移学習のベースと微調整では解像度が異なる。

用いたネットワークは過去のコンテストで優秀だったものをベースとしている。

- VGG-CNN-M-2048

“Visualizing and Understanding Convolutional Networks” Matthew D. Zeiler and Rob Fergus
引用数2631

- AlexNet

“ImageNet Classification with Deep Convolutional Neural Networks” Alex Krizhevsky
引用数17181

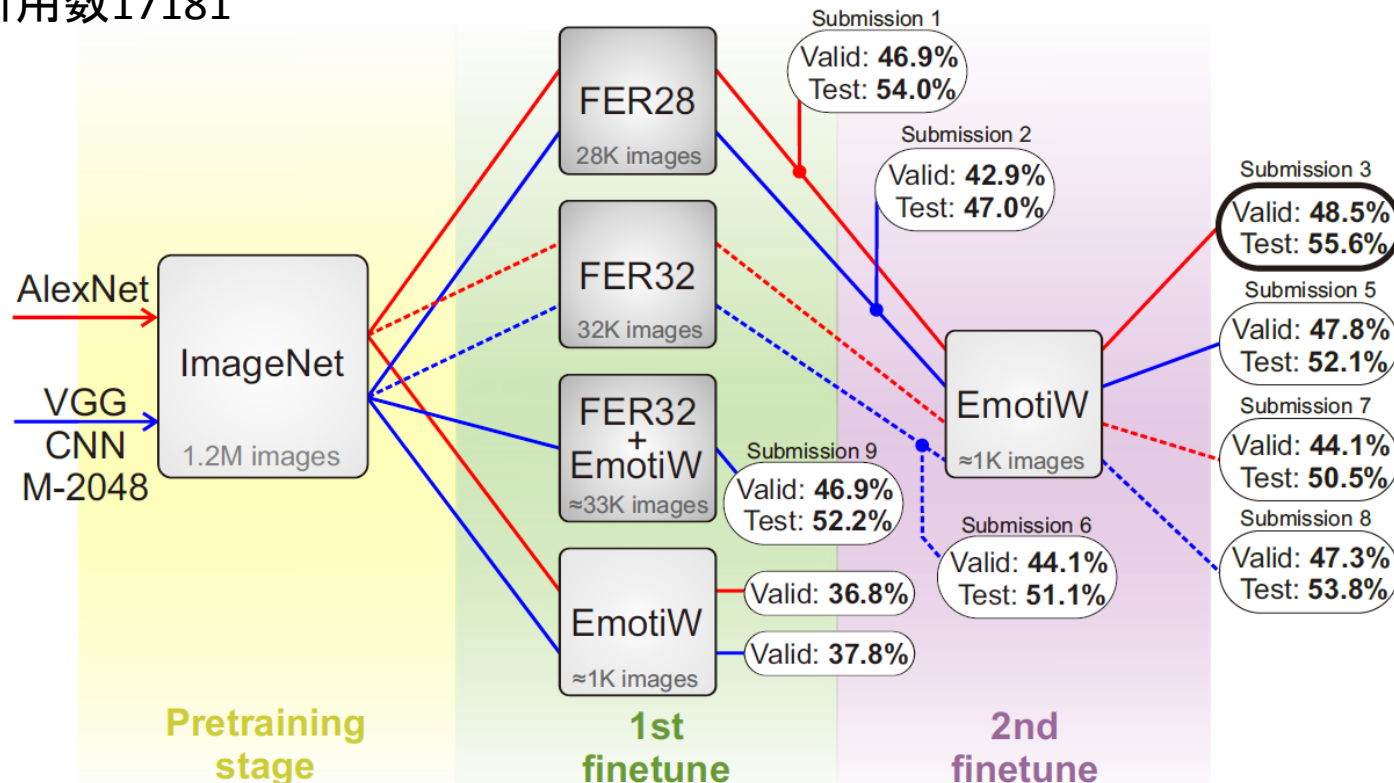
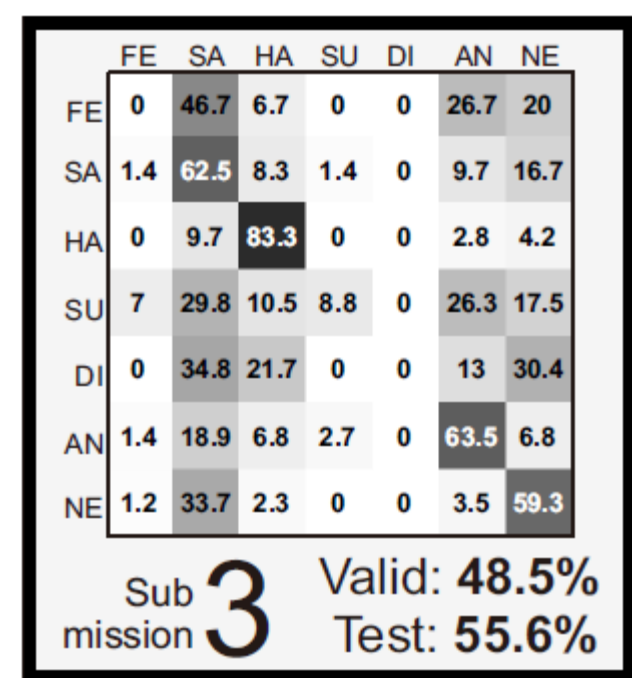
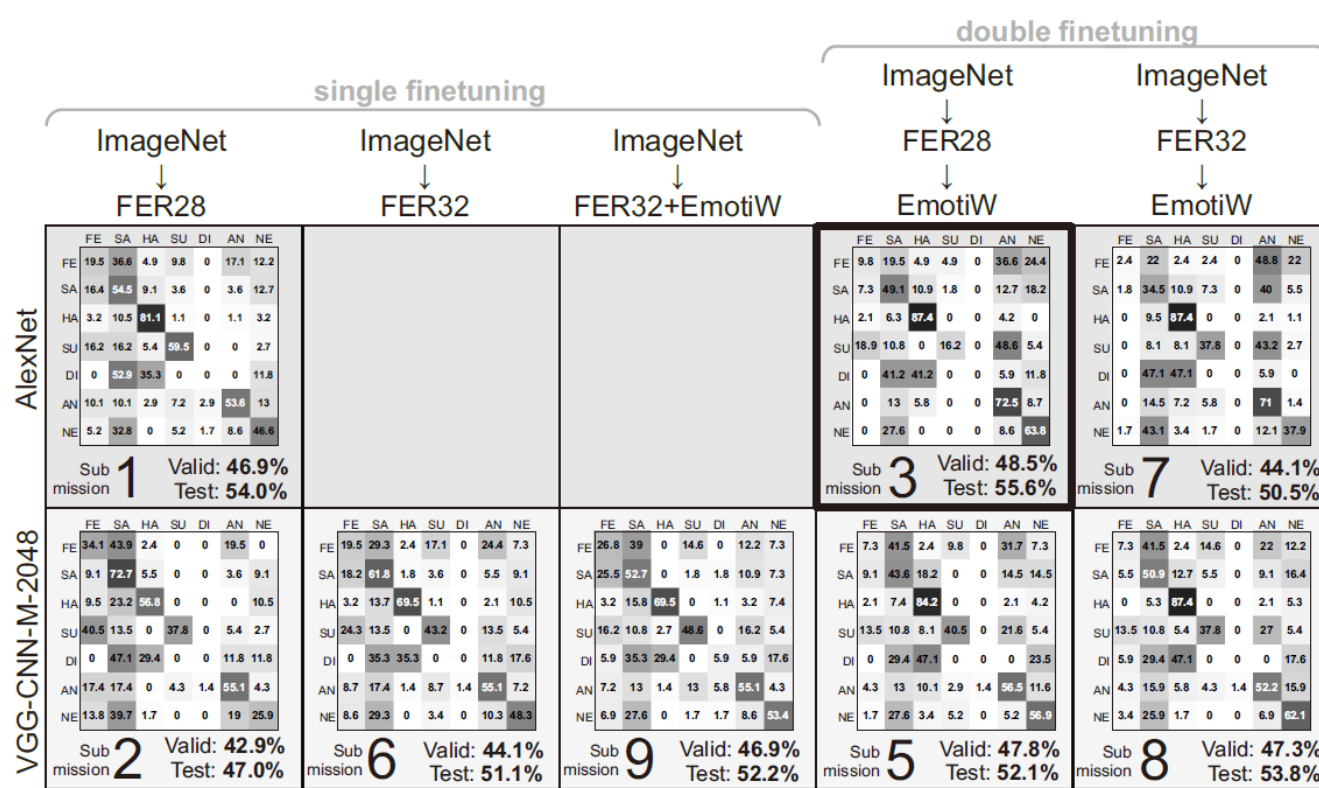


Figure 3: Flowchart of the different fine-tuning combinations used. Our 3rd submission exhibited the best performance.

転移学習を行うことで認識性能が向上した。

FER32よりFER28の方が性能がいいのは本番データとの解像度の違いなどの影響か。



恐怖と嫌悪は難しい。
幸福は83%正答。

Figure 4: Comparison between the confusion matrices of different models for the EmotiW test set. Submission 4 was omitted due to mistakes in training the model. Other omissions were due to lack of time to prepare the models.

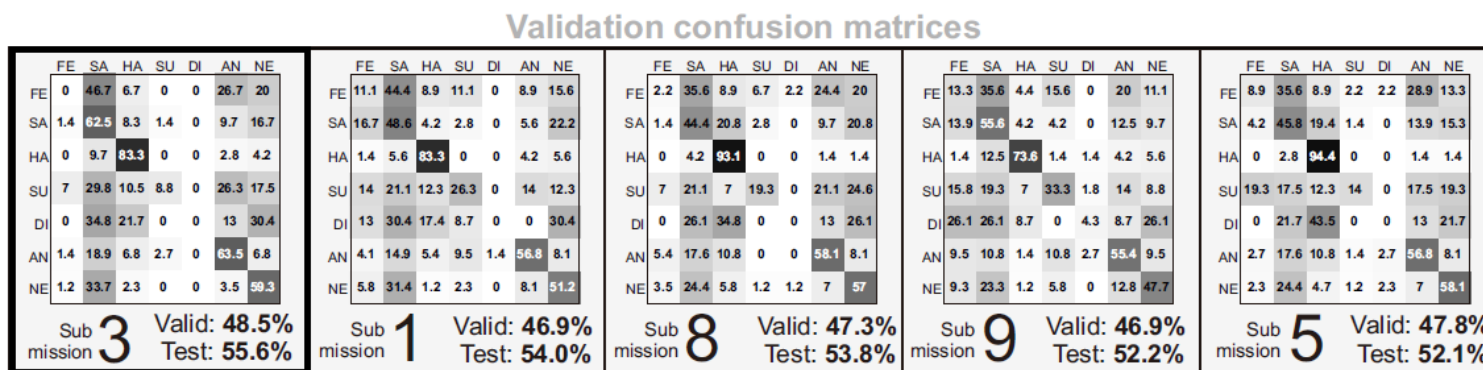


Figure 5: Confusion matrices generated from the EmotiW validation set for our 5 best submissions.

Happyを多少弱めて、
Sadを強化することで
全体として向上

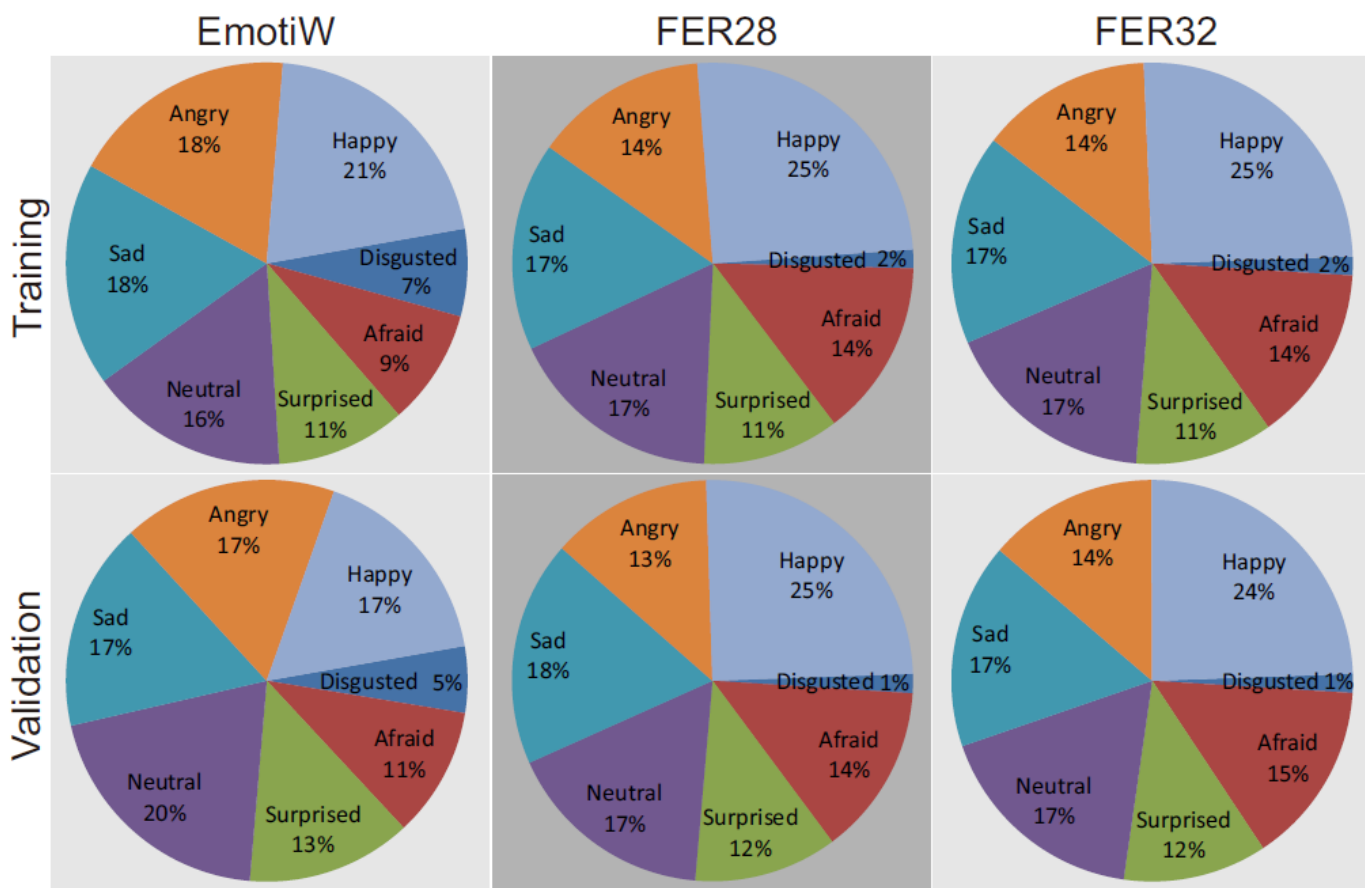


Figure 6: Distributions of the 7 classes on the EmotiW, FER28 and FER32 training and validation sets.

学習、評価に用いたデータどちらも偏りのあるデータセット。

怒っている人の写真・ムービーを撮る機会は少ない。

運営が出している感情の“正解ラベルが間違っている”可能性も

効率的に感情ラベルのついた画像・動画のデータベースを作るのも重要なタスク

- Collecting Large, Richly Annotated Facial-Expression Databases from Movies
- JOURNAL OF LATEX CLASS FILES, VOL. 6, NO. 1, JANUARY 2007
- Abhinav Dhall,

映画から半自動で感情ラベルをつけた画像・動画データベースを生成。
他のデータベースとの比較を行う。(含まれている被験者の年齢、ラベルの複雑さなど)

参考: 顔画像認識に使える10のデータセット

<https://qiita.com/Hironsan/items/de739575cdd866226676>

Labeled Faces in the Wild

名前	Labeled Faces in the Wild
カラー画像	Yes
画像サイズ	150x150
ユニークな人数	13000人以上
画像数/人	1680人は2枚以上の画像が存在

Faces in the Wild

名前	Faces in the Wild
カラー画像	Yes
画像サイズ	86x86
ユニークな人数	15000弱
画像数/人	1枚以上

Face Recognition Database

名前	Face Recognition Database
カラー画像	Yes
画像サイズ	大きめ
ユニークな人数	10
画像数/人	5枚くらい
その他	3D head modelの画像は大量にある(324/人)

Sheffield Face Database

名前	Sheffield Face Database
カラー画像	No
画像サイズ	92x112, 248x258
ユニークな人数	20
画像数/人	20~40

ユーザー生成コンテンツの爆発的増加

- **Web、特に画像投稿・共有サービスでは...**

– Flickr	画像100億枚	2015年5月時点
– Instagram	画像400億枚	2015年9月時点
– Facebook	画像2500億枚	2013年9月時点



Pohutukawa blooms this time of the year in New Zealand. As the flowers fall, the ground underneath the trees look spectacular.

大量に収集可能：
画像と関連する文章の対

まとめ

- ・マイクロソフトCognitive Serviceなどで感情認識技術への関心は高まっており、今後更なる実用化も進むと考えられる。
- ・日銀総帥の会見の表情分析はノイズが偶然一致したとも取れるが、アイディアは参考になる
- ・転移学習を用いたAI開発で感情認識コンテストで優秀な成績を収めた。
- ・感情のラベルつきデータの収集というひとつの分野もある。