Text and Image Synergy with Feature Cross Technique for Gender Identification

CLEF/PAN 2018 Author Profiling Task

September 10, 2018

Takumi Takahashi, Takuji Tahara,
Koki Nagatani, Yasuhide Miura,
Tomoki Taniguchi, and Tomoko Ohkuma

Fuji Xerox Co., Ltd.

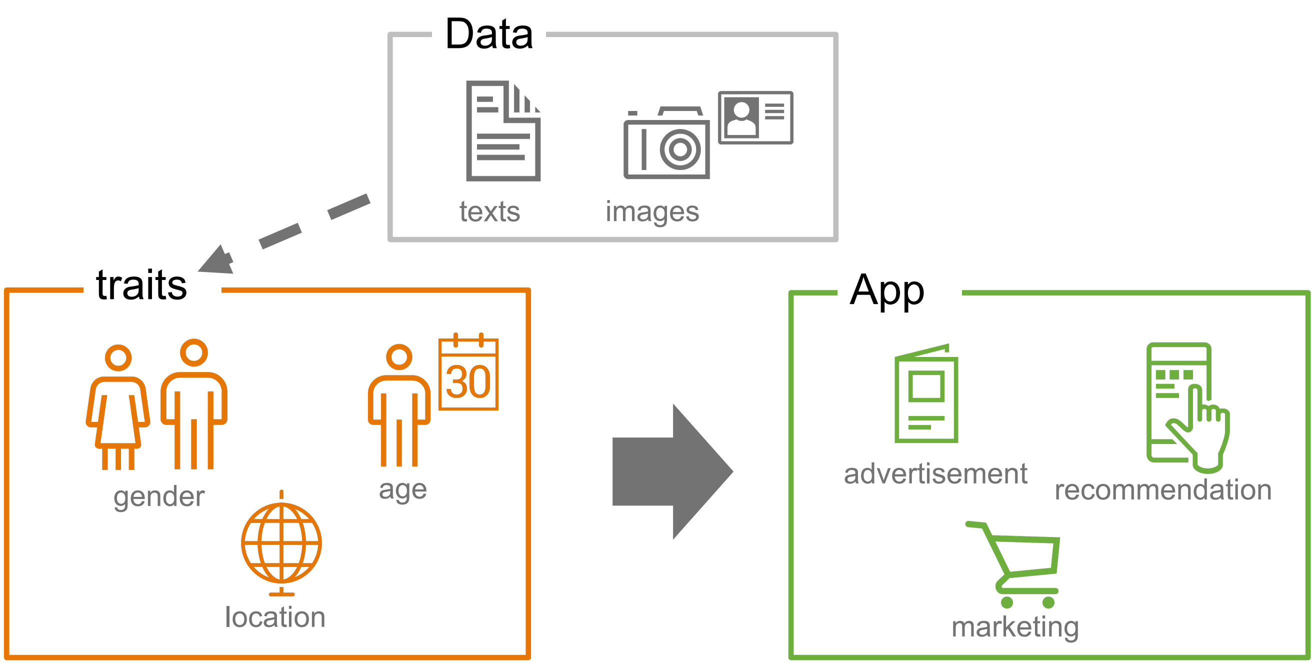
Outlines

- Introduction
- PAN 2018 Author Profiling Task
- Related Work
- Our Motivation
- Proposed Model
- Experiment
- Result
- Discussion
- Conclusion & Future Works

1. Introduction

■ Author profile traits on social media:

- Author profile traits can be applied to some app
  - traits: age, gender, location, …
  - App: advertisement, recommendation, marketing, …etc



■ Issues:

- Author profile traits are not explicitly described on social media.
  - This causes difficulty to utilize author profile traits on app

2. PAN 2018 Author Profiling Task

Gender identification from Tweets:

- Gender identification:
  - Binary classification from Tweets (male/female)

- Target languages:
  - Arabic, English, Spanish

- Datasets:
  - Text data contains 100 Tweets for each user
  - Image data contains 10 images for each user

New dataset in PAN 2018

	Users	Tweets	Images
Arabic	1,500	150,000	15,000
English	3,000	300,000	30,000
Spanish	3,000	300,000	30,000

3. Related Work (1)

■ Strong models at PAN 2017:

- Traditional machine learning approaches successfully performed
  - Linear SVM with character 3- to 5-grams and word 1- to 2-grams features (Basile et al., 2017)
  - Exploring many approaches and employing logistic regression (Martinc et al., 2017)
  - Micro TC: generic framework for text classification (Tellez et al., 2017)

■ Deep Neural Network approaches at PAN 2017:

- DNN approaches were also presented
  - Bi-directional GRU with attention for word + CNN for character (Miura et al., 2017)
  - CNN with convolutional filters of different sizes (Sierra et al., 2017)

➡ In PAN 2017, DNN could not outperform traditional ML models

3. Related Work (2)

■ Author profiling tasks outside of PAN:

・Combining both texts and images in neural network
  - Prediction user’s traits (gender, age, political orientation, and location)
  - The model that utilized both texts and images showed state-of-the-art performances (Vijayaraghavan et al., 2017)

■ Expectation:

・Utilizing not only texts but images would be effective for author profiling

4. Our Motivation

■ Deep Neural Network (DNN):

- In PAN 2017: DNN approach showed 4th ranking (Miura et al., 2017)

■ Main approaches at PAN 2017:

- Traditional machine learning approaches successfully performed
  - SVM, Random Forest, Logistic Regression, …
  - Uni-gram, Bi-gram features were often employed

■ Unveiling images:

- PAN 2018 unveiled images to identify user's gender
  - 10 images are prepared for each user
  - Many successful models exist in CV tasks (AlexNet, VGG16, ResNet)

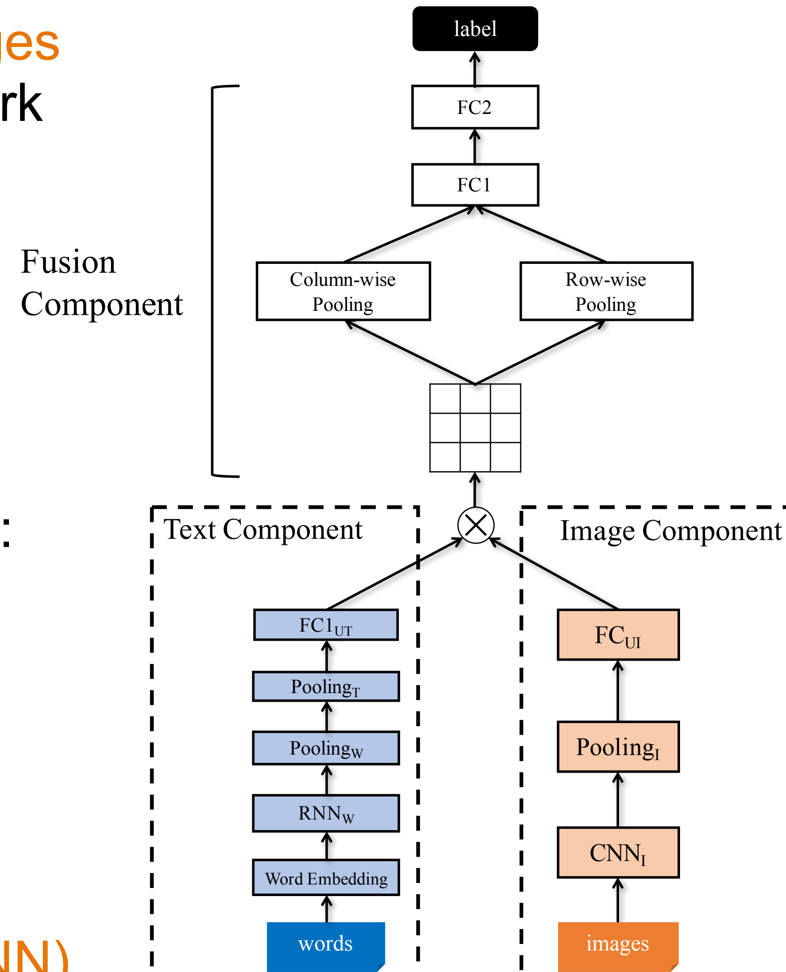
➡ Performances will be enhanced combining texts with images in DNN

5. Proposed Model

■ Core idea

- Leverage the synergy of both texts and images with feature cross technique in neural network
- Relationship between both features are computed by direct-product

→ Inspired by (Santos et al., 2016) for QA



■ Major components

The model is constructed of three components:

1. Text Component:
2. Image Component:
3. Fusion Component

➡ Text Image Fusion Neural Network (TIFNN)

5-1. Text Component

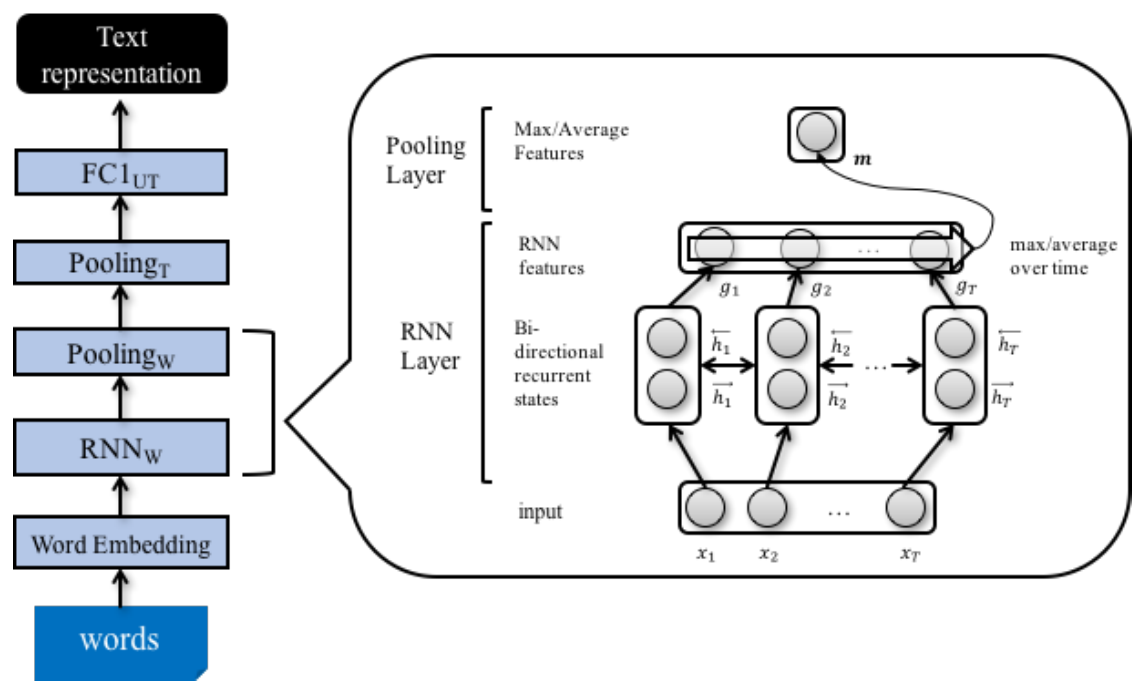
Purpose of the component:

- Encoding text representation from user's Tweets
- Integrating 100 Tweets for each user into a representation

Model composition:

- RNN_W: The layer is constructed of bi-directional GRU
- $Pooling_W$: Integrating words in a tweet (word-level pooling)
- $Pooling_T$: Integrating tweets in a user (Tweet-level pooling)

RNN_W: handles sentence word by word (each time step t)



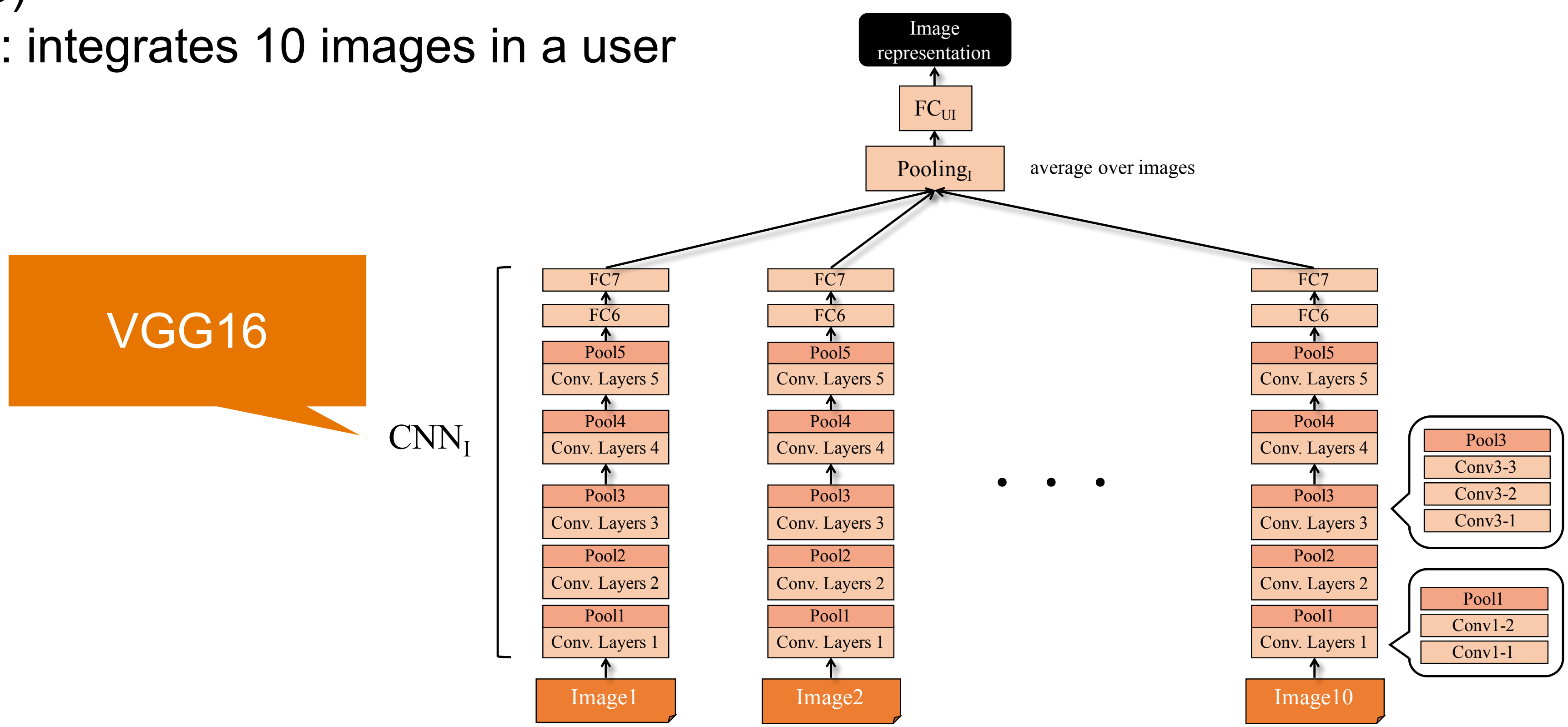
5-2. Image Component

■ Purpose of the component:

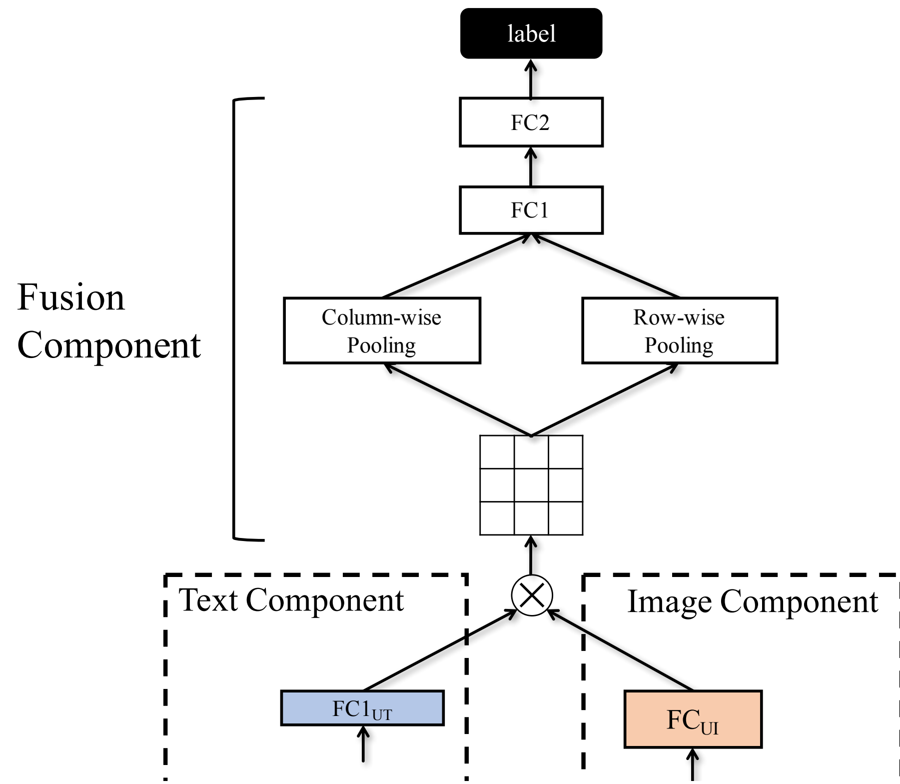
- Encoding image representation from each user
- Integrating 10 images for each user into a representation

■ Model composition:

- $\mathrm{CNN_I}$: 13 convolutional layers, 5 pooling layers, 2 fully connected layers (VGG16)
- $\mathrm{Pooling_I}$: integrates 10 images in a user



5-3. Fusion Component



■ Purpose of the component:

- Leveraging synergy of both texts and images by feature cross technique
- Finally, the model classifies user’s gender using combined feature

■ Model composition:

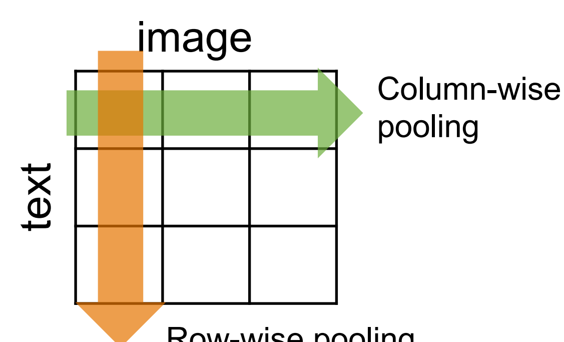
- direct-product: captures the relationship between texts and images

$$\boldsymbol{G} = \boldsymbol{r}_{txt} \otimes \boldsymbol{r}_{img}$$

- Column-wise pooling: finds out the most relevant image element with respect to text representation

$$[g_{txt}]_j = \max_{1 \le l \le L} [G_{j,l}]$$

$$[g_{img}]_j = \max_{1 \le m \le M} [G_{m,j}]$$



6. Experiment

■ Dataset:

- PAN 2018 Author Profiling Task Corpus:
  - divided this corpus into $train_8$, dev_1, and $test_1$ with a gender ratio 1:1

	$train_8$	dev_1	$test_1$	Full size
Arabic	1,200	150	150	1,500
English	2,400	300	300	3,000
Spanish	2,400	300	300	3,000

■ Streaming Tweets:

- Collected Tweets to pre-train the word embedding matrix $\boldsymbol{E_w}$ from Twitter by Twitter Streaming APIs
  - During the period of March-May 2017
  - Remove Retweets
  - Delete Tweets posted by bots

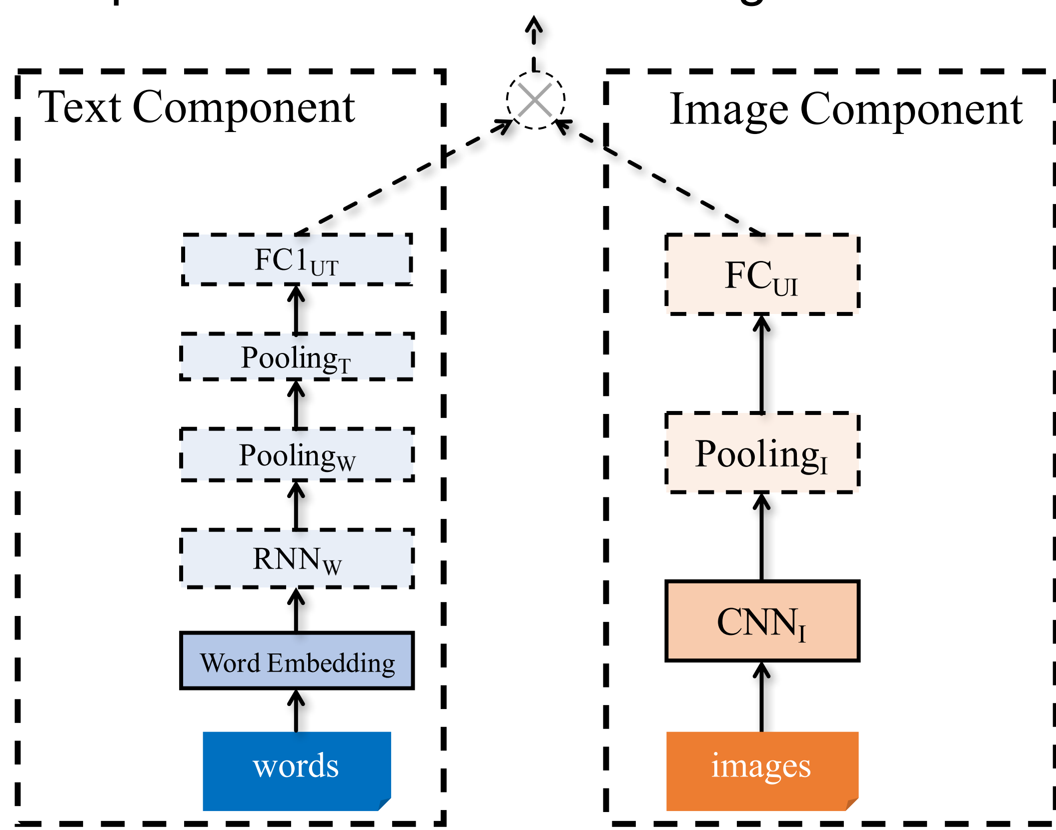
	# of Tweets
Arabic	2.46M
English	10.72M
Spanish	3.17M

6-1. Training Procedures (1)

■ Pre-train word embedding & VGG16

- Initialization of word embeddings:
  - Utilized fastText with the skip-gram algorithm to pre-train word embedding (Bojanowski et al., 2016)

- Initialization of $\mathrm{CNN_I}$
  - $\mathrm{CNN_I}$ is initialized with parameters of pre-trained VGG16 on ImageNet



6-1. Training Procedures (2)

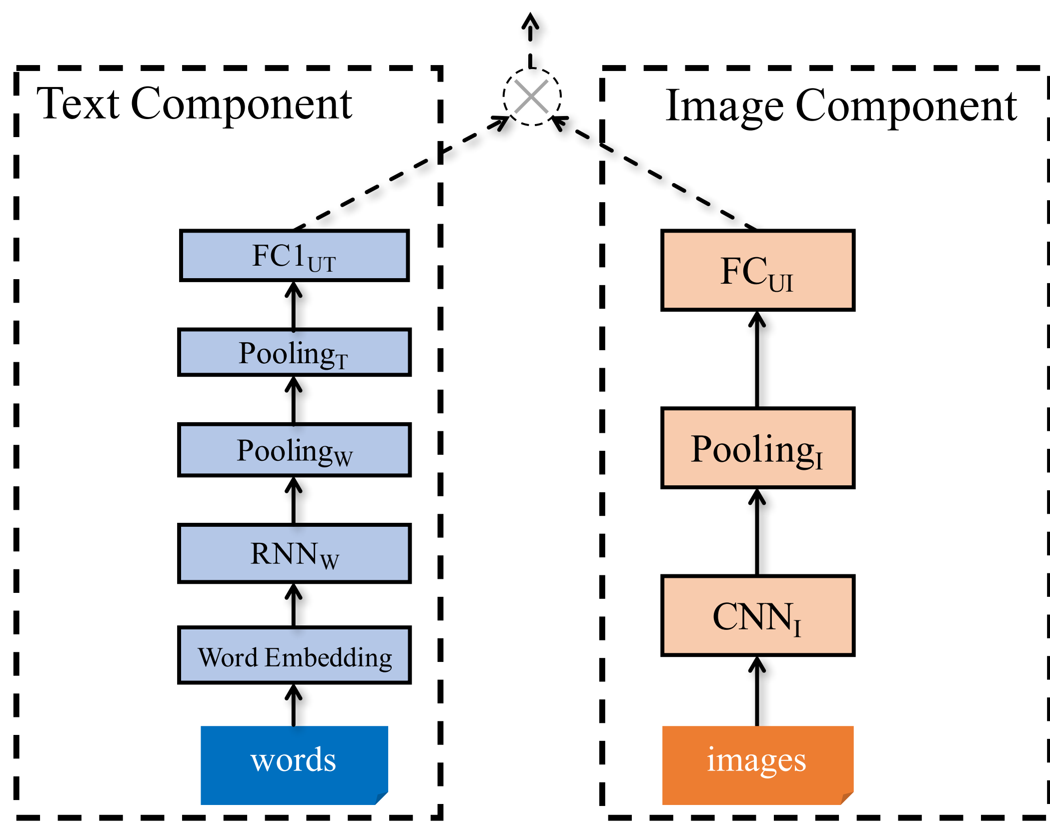
■ Component-wise training:

- Text component:
  - Text component is trained using $train_8$ and dev_1

- Image component:
  - Image component is trained using $train_8$ and dev_1

NOTE:

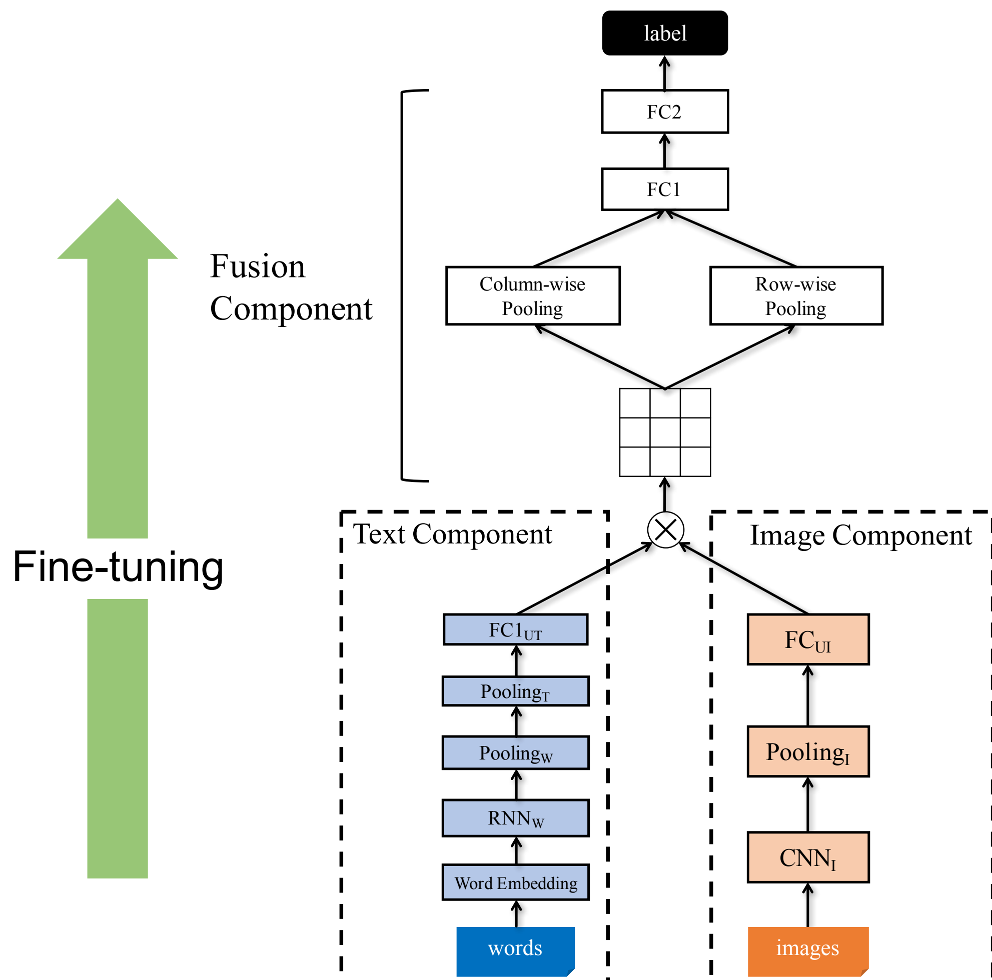
Each component is trained without <u>fusion component</u>!!



6-1. Training Procedures (3)

■ TIFNN training:

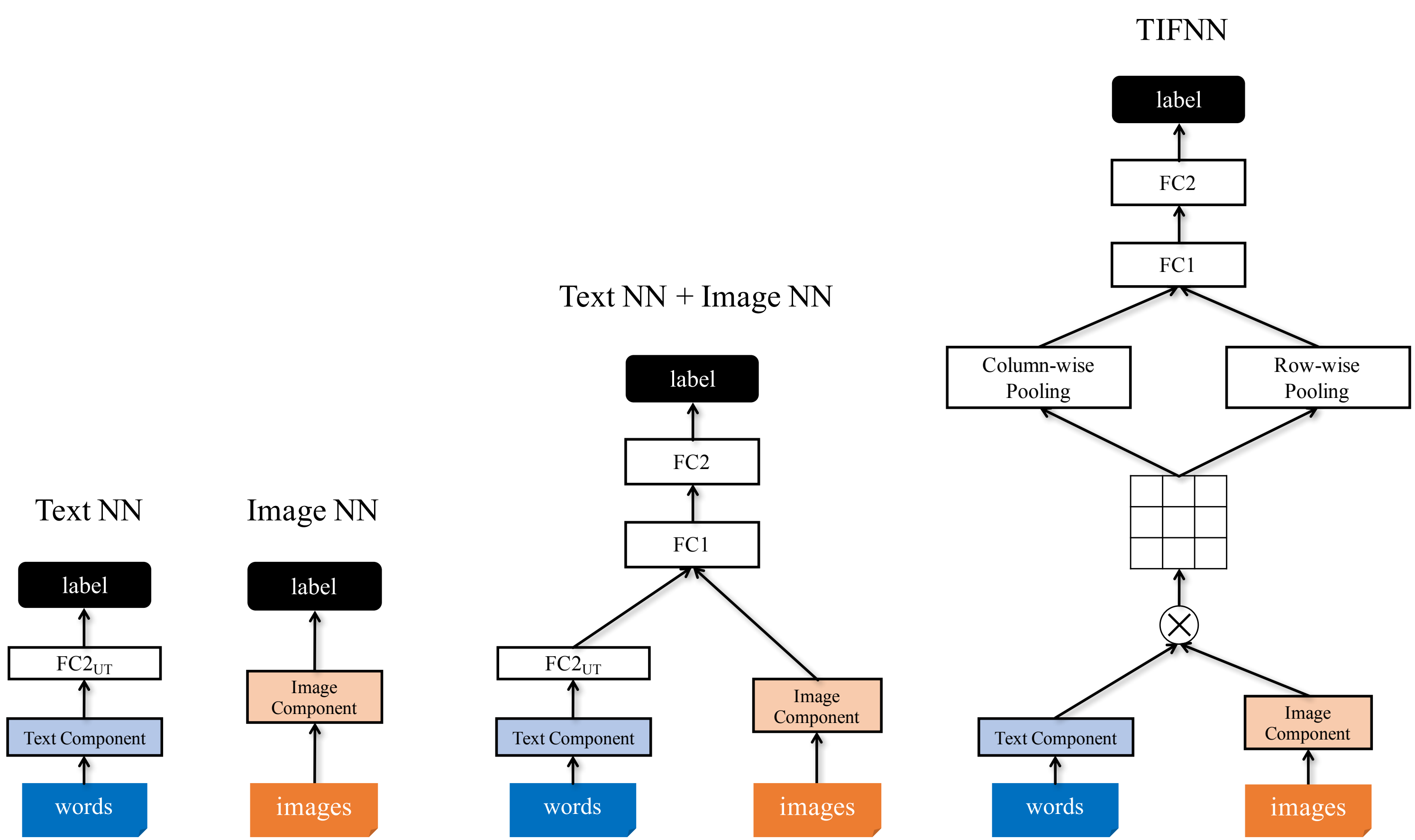
- All of TIFNN parameters except final FC layers are initialized with parameters of the pre-trained components
  → The entire model is trained by fine-tuning using train_8 and dev_1



6-2. Comparison Models

Comparison Models:

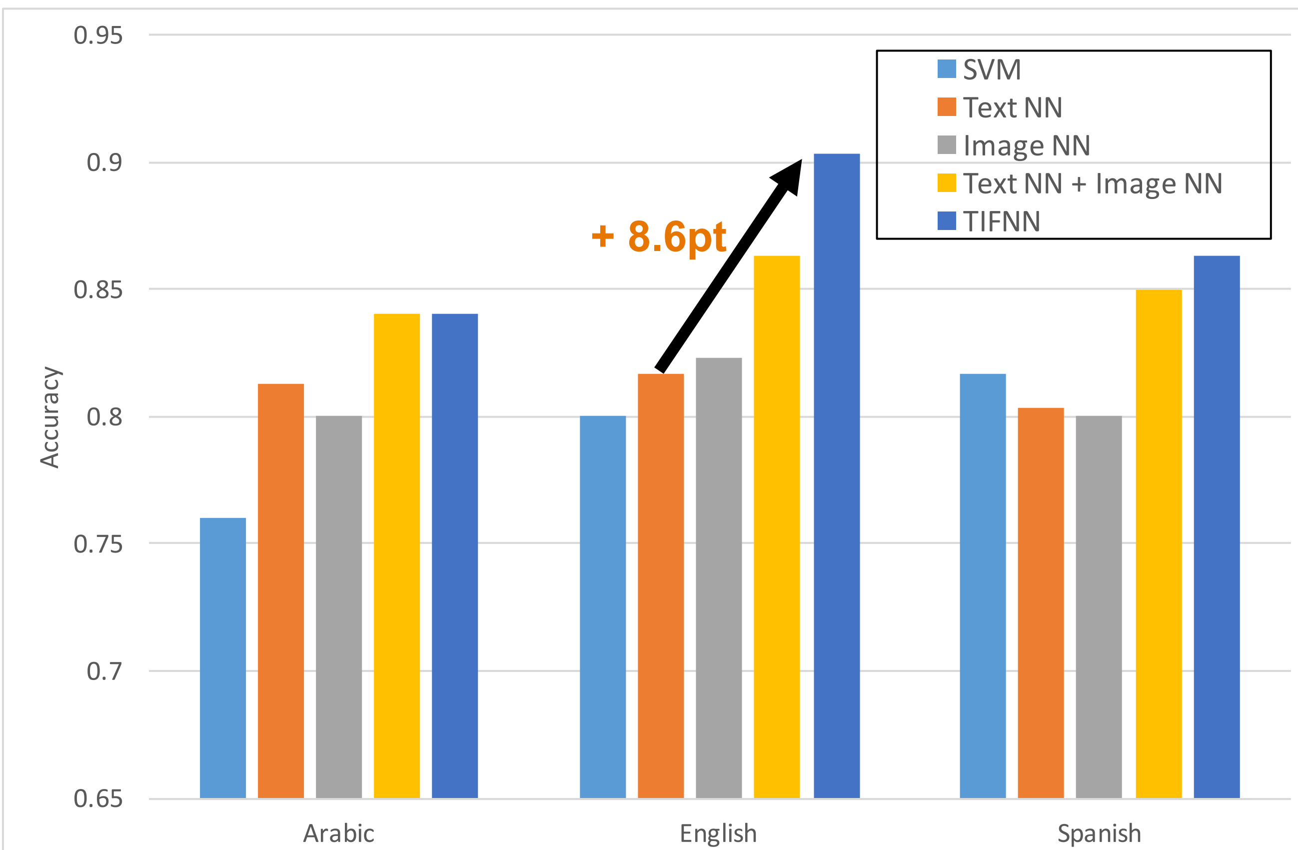
- **SVM**: SVM using TF-IDF uni-gram features; strong baseline
- **Text NN**: Text component and a fully connected layer
- **Image NN**: Image component
- **Text NN + Image NN**: Combines both NNs without fusion component



7. Result (In-house Experiment)

■ In-house experiment:

- **Text NN** and **Image NN** achieved accuracies of 80.0-82.3%
- **TIFNN** drastically improved the accuracies: + 2.7-8.6pt !
  - Significantly improved for English
- **TIFNN** also outperformed **Text NN + Image NN**

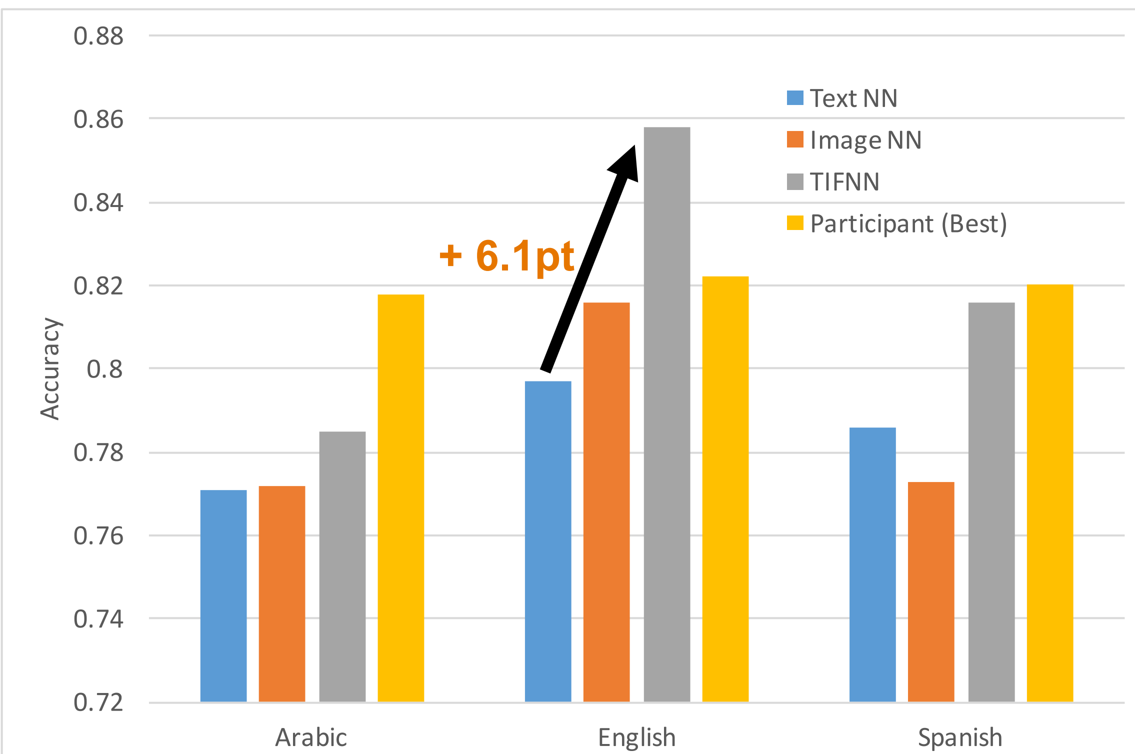


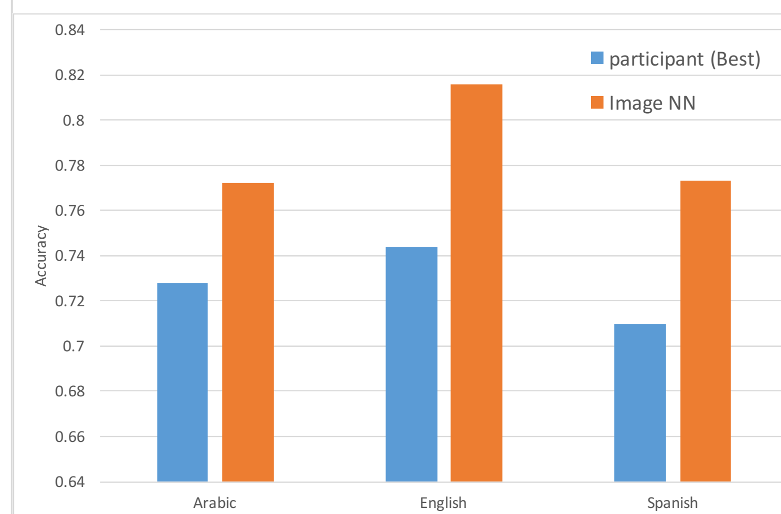
7. Result (Submission Run)

Submission run:

- **TIFNN** had better accuracies compared with individual models (1.3-6.1pt)
  - The model had lower accuracies compared with In-house experiment
  → Perhaps overfitting

- Image NN significantly outperformed other systems
- **Ranked 1st** in entire participants





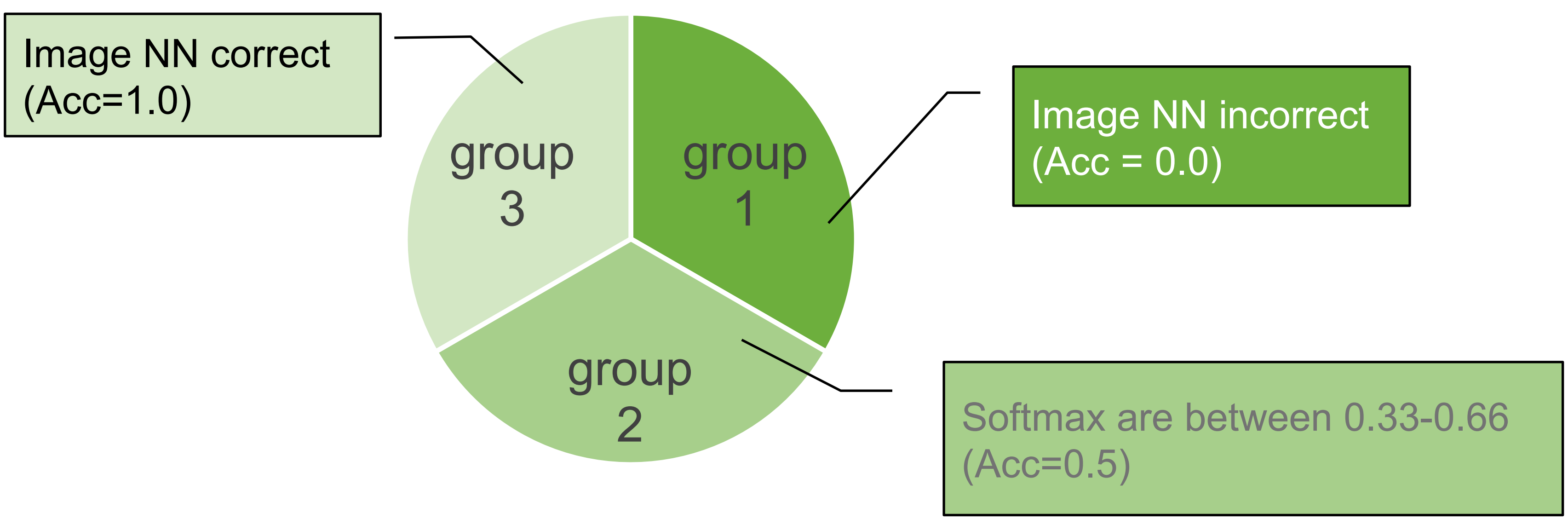
8. Gender Identification by Human (1)

■ Correlation between Human and Image NN:

・Image NN showed superior performances in this task
  - How much accuracies can humans identify user's gender from images?
  → Investigating the correlation between human and Image NN

■ Categorizing target users:

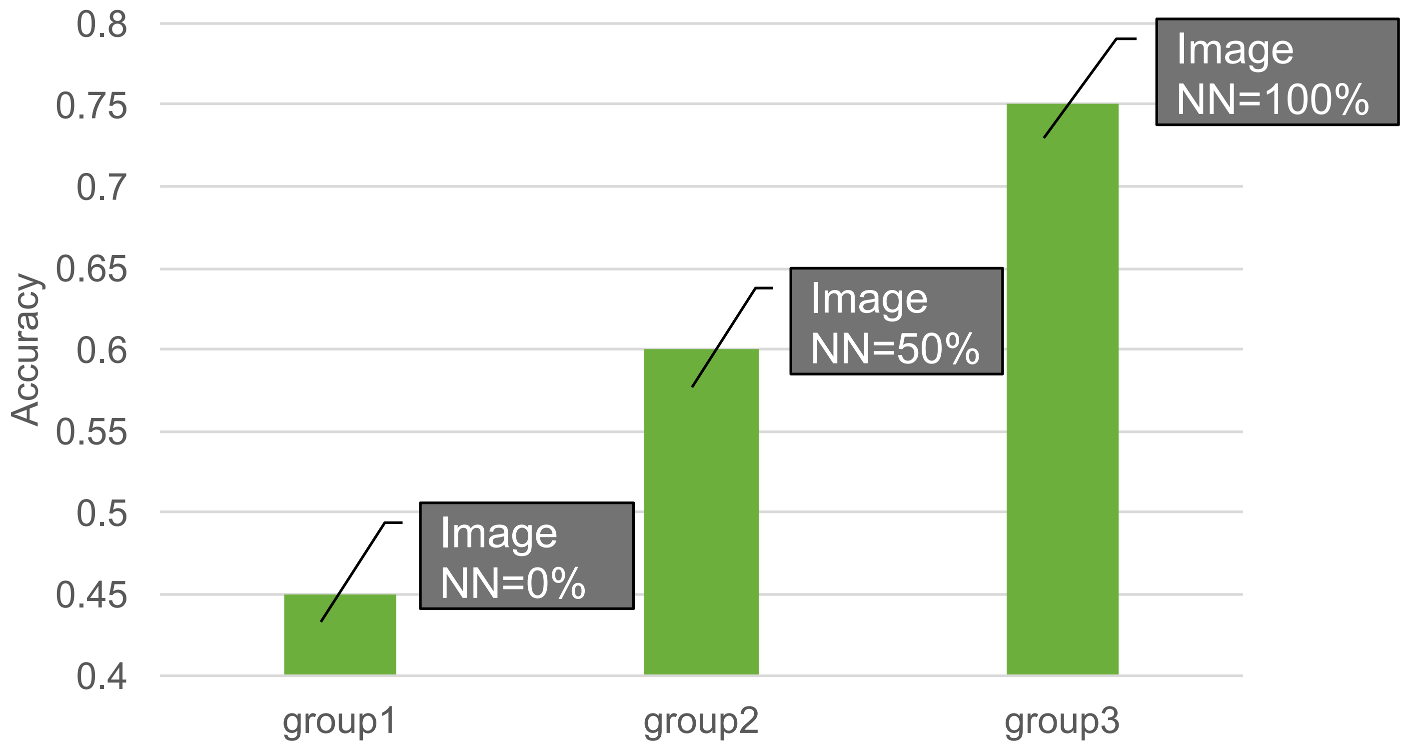
・Target users were divided into 3 types of category:



8. Gender Identification by Human (2)

■ Experimental result:

・The trend is the same between human and Image NN
- group1: Human can identify user’s gender with 45% accuracy
- group2: The accuracy is better 10% than Image NN
- group3: The accuracy drops 25% compared with Image NN



Average accuracy: 60%

9. Conclusion & Future Works

■ Conclusion:

- Proposed Text Image Fusion Neural Network (TIFNN) for gender identification
  - Components:
    - Text component
    - Image component
    - Fusion component
- Improvement compared with individual models
  - In-house experiment: + 2.7-8.6pt for each language
  - Submission run: + 1.3-6.1pt → **Ranked 1st** in entire participants

■ Future Works:

- Analyzing how the proposed model interacts with texts and images
  - Understanding this interaction makes it possible to improve TIFNN

Thank you !!