YouTube View Count Prediction

Team 6

Suho Park
Chaejin Lim
Jaejoon Yoo
Jeongyeop Seo
Jihyun Lee

Contents

01 Background

02 Contribution

03 Model Design

04 Experiment Setting

05 Evaluation

06 Conclusion

01

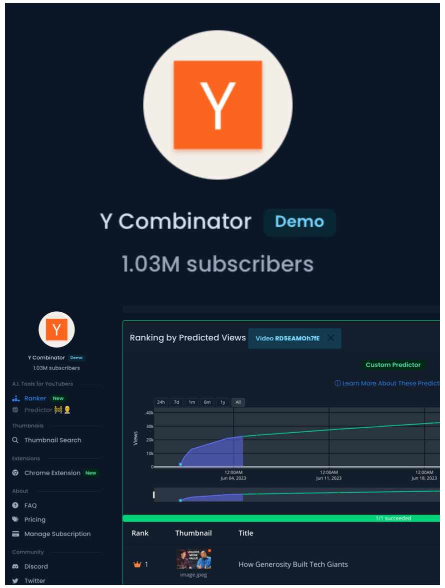
Background

Existing Work

Prediction of future view counts of a specific YouTuber's uploaded videos

Prediction of view counts utilizing numerical features (meta)

Study on feature importance of thumbnail and meta-level features



Problems of Existing Work

Limited to certain YouTuber

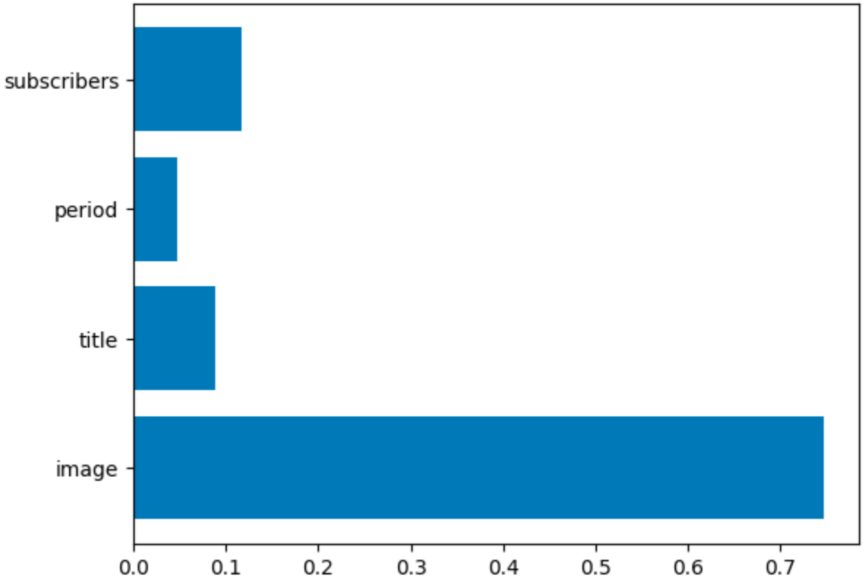
No help in deciding thumbnail or title

Not utilizing thumbnail features

Feature Importance

Use XGBoost to measure

how much the corresponding feature affects prediction



02

Contribution

Our Goal

Predict the view counts of a YouTube Mukbang video utilizing thumbnail features

Main Contribution

Crawled and used our own dataset to solve the problem

Designed an optimal architecture to utilize features in balance including thumbnail features.

Successfully predict view counts of a YouTube mukbang video prior to upload

HTTPSnet

Hierarchical prediction model with Thumbnail, Title, Period and Subscribers

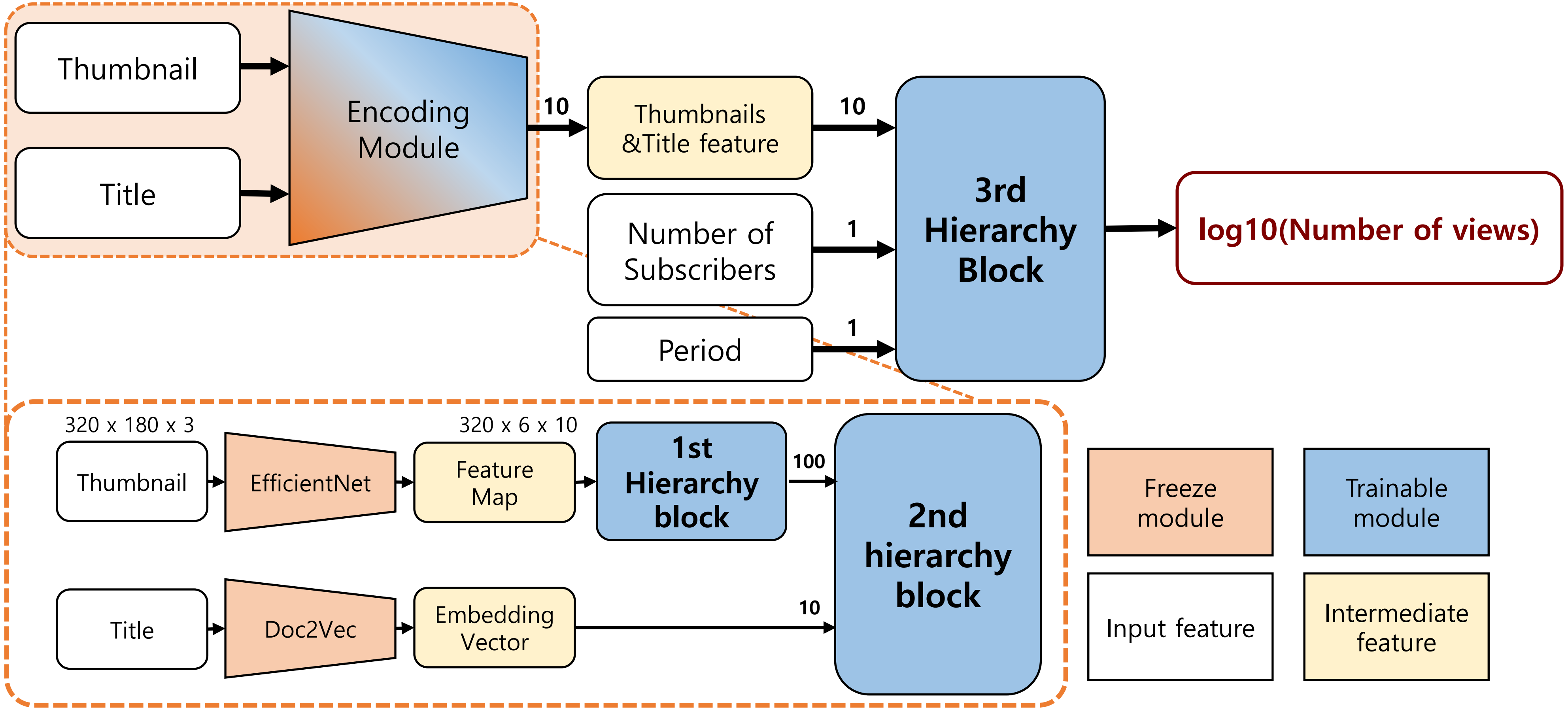
03

Model Design

Overall architecture of our HTTPSnet

We designed **hierarchical architecture** model for view count prediction

EfficientNet and **Doc2Vec** are utilized for **representing thumbnail and title as a vector**



Why hierarchical architecture?

There is a serious imbalance in the size of the input vectors.

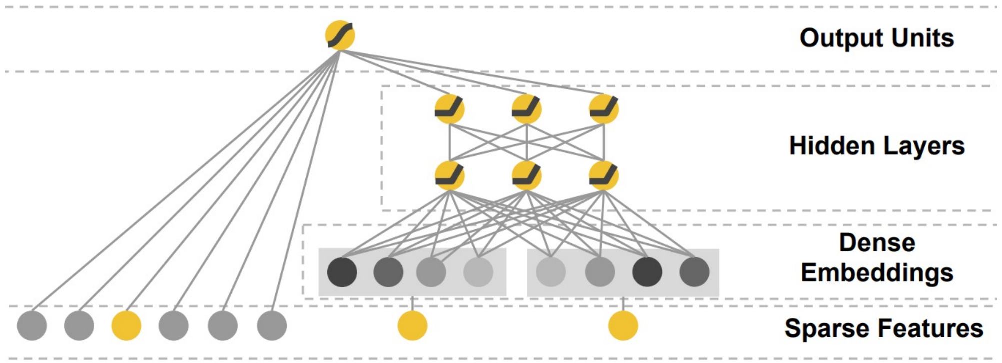
Input type	Shape of vector
Thumbnail	(320,6,10)
Title	(10,1)
Upload period	(1,1)
Number of subscribers	(1,1)

Too large image vector can make other input vectors become faint

Wide & Deep Learning for Recommender Systems

Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah

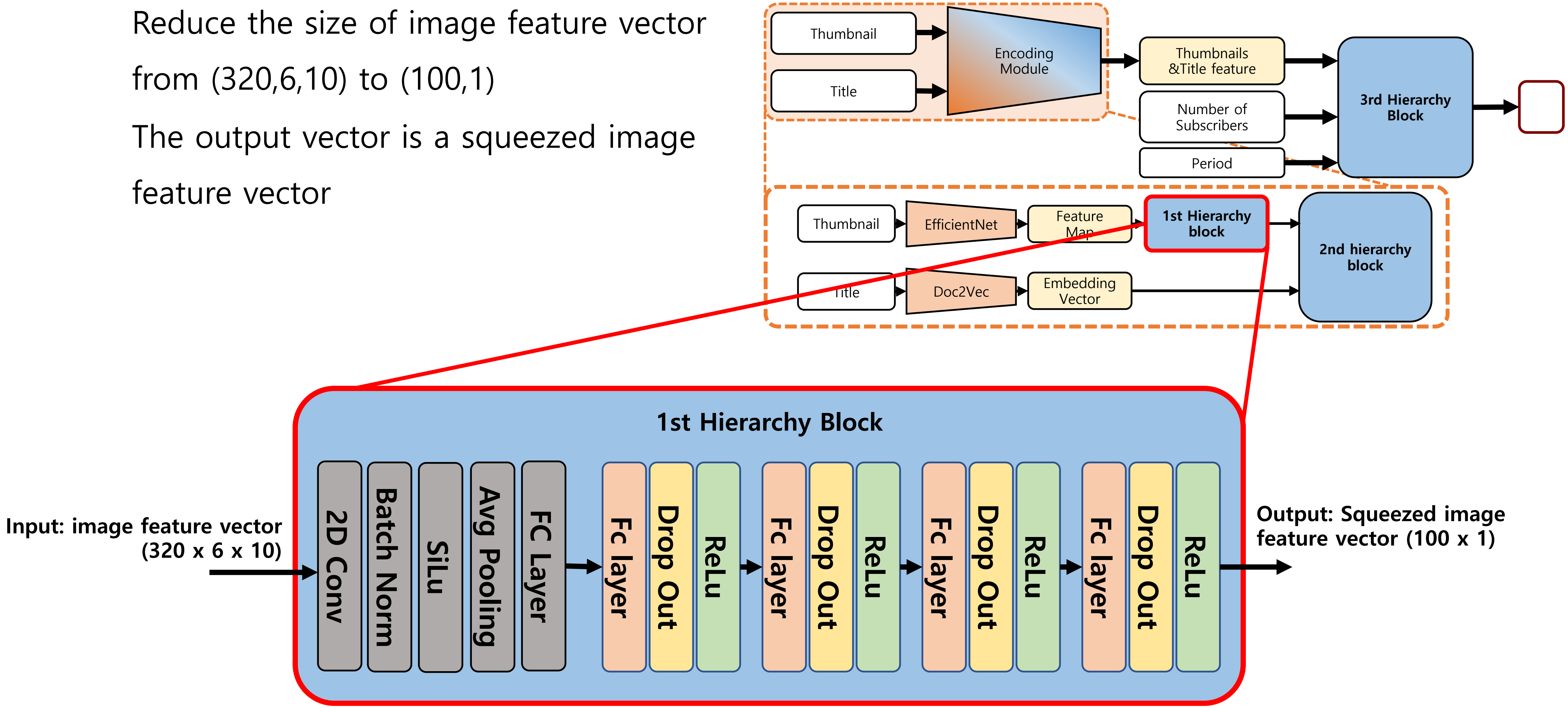
Google Inc.*



1st Hierarchy Block

Reduce the size of image feature vector from (320,6,10) to (100,1)

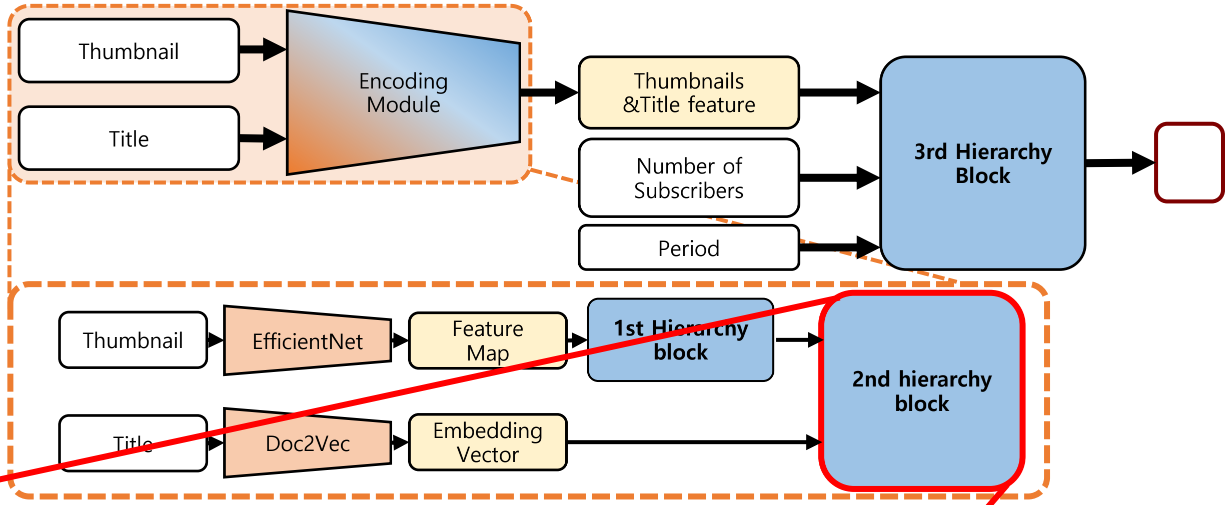
The output vector is a squeezed image feature vector

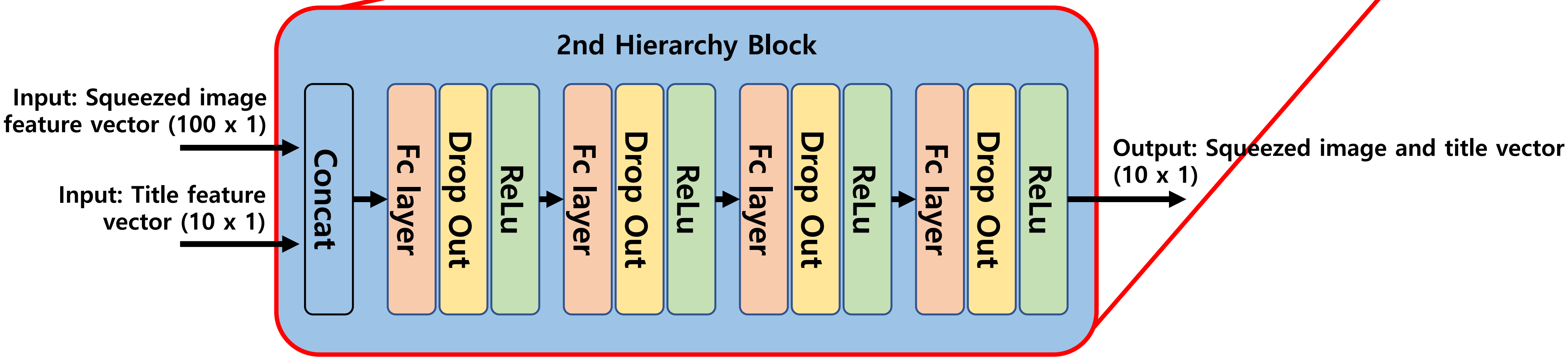


2nd Hierarchy Block

Squeezed image feature vector and title feature vector are concatenated

The output vector is a squeezed image and title feature vector with shape(10,1)

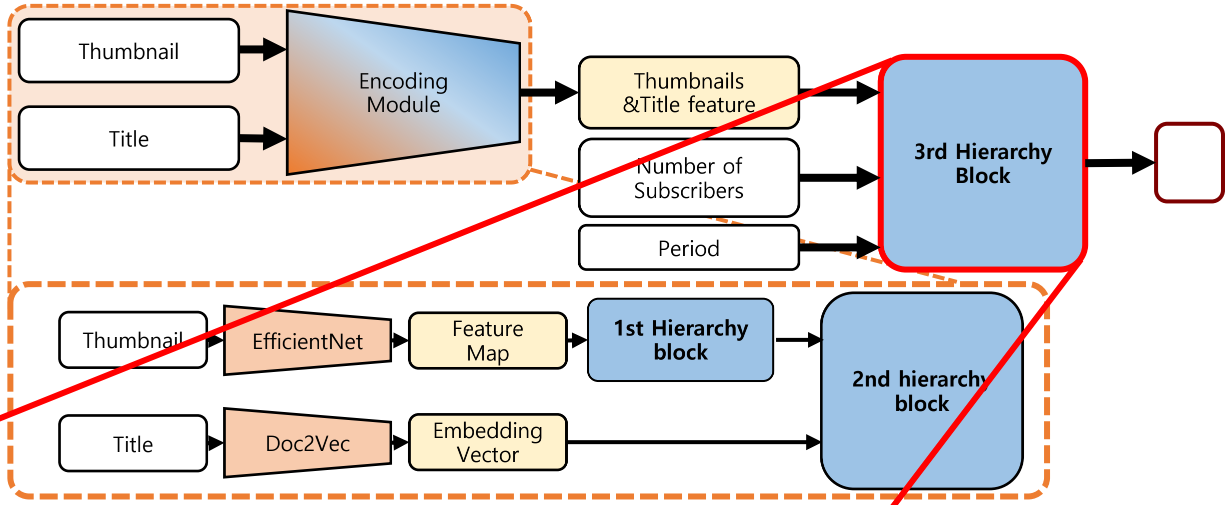


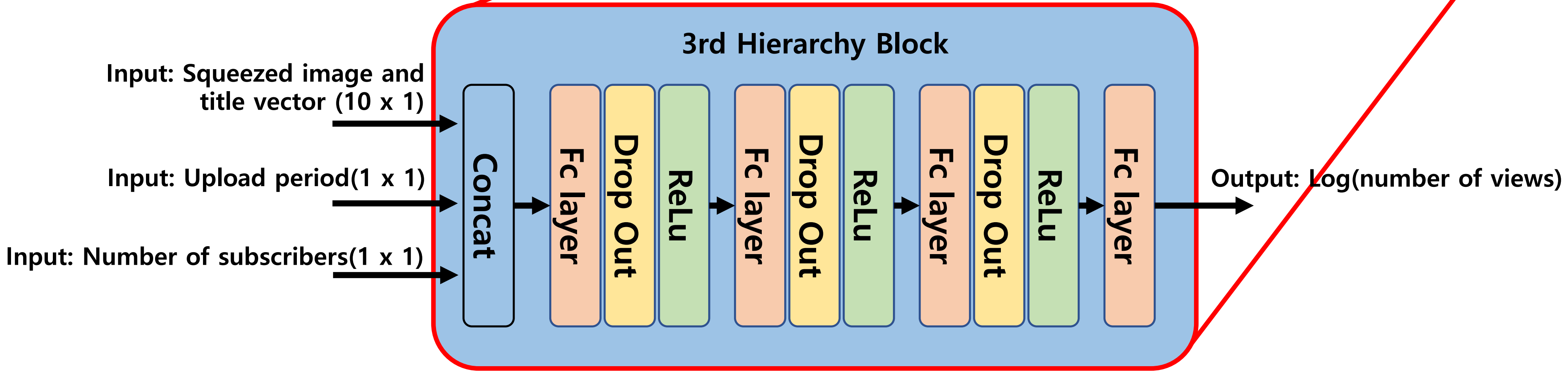


3rd Hierarchy Block

The final stage of our model

The output value is prediction of logarithm of view counts





The effect of hierarchical model

The model can learn with balanced size of inputs

Input type	Shape of vector
Thumbnail	(320,6,10)
Title	(10,1)
Upload period	(1,1)
Number of subscribers	(1,1)

After
hierarchy blocks

Input type	Shape of vector
Thumbnail and Title	(10,1)
Upload period	(1,1)
Number of subscribers	(1,1)

More balanced input size than before!

04

Experiment Setting

Data Collection

Youtube V3 API: keyword search, related video search

Collected Features: publish time(⇢period day), subscriber count, views, thumbnail, video title, …

Filter out non-Mukbang videos: theme category provided in video meta data

Total 15,287 samples

Full Dataset

Train Dataset: 12,382 (81%)

Validation Dataset: 1,376 (9%)

Test Dataset: 1,529 (10%)

	video_id	publish_time	publish_date	channel_id	title	views	period_day	channel_title	subscriber_count
0	9Qjs1PmepkM	13:00:12	2021-11-04	UCcfKn5ex1g8zgK4eYtbReoA	3시간동안 털어왔다...🥤 편의점 음식 디저트 먹방 ASMR MUKBANG | 초콜릿 ...	3277204	551.0	코지 COSY	457000
1	WElm5p5wP2Q	11:30:55	2023-05-07	UCQgrXvcEplOcH8Z7F7T8WuA	Mukbang mix.#fish#ftuits#banana leave#mango #...	1211	2.0	Mina Kim	158
2	7Bt7P-LCVGM	10:56:43	2023-02-23	UCGxgxnbz-KnZylyXkjS3VAQ	Eating fruit cake bubble milk Mukbang	2904	72.0	Eaten Show	71500
3	SoQ0JD0pHt0	22:05:45	2023-02-18	UCeTLzXqW8DypCLnu9NHYcPA	watch this video if you are stressed	1790	79.0	AxolotlPlayz	28
4	hreGA3Xr8QE	13:15:45	2020-06-22	UCjmO0H7AV7FfOIExXB3VJ9Q	MOIST & SOFT BANANA CAKE (NEGOSYO IDEA)	2317933	1048.0	TiNa TaMz	25300
...	...	...	...	...	...	...	...	...	...

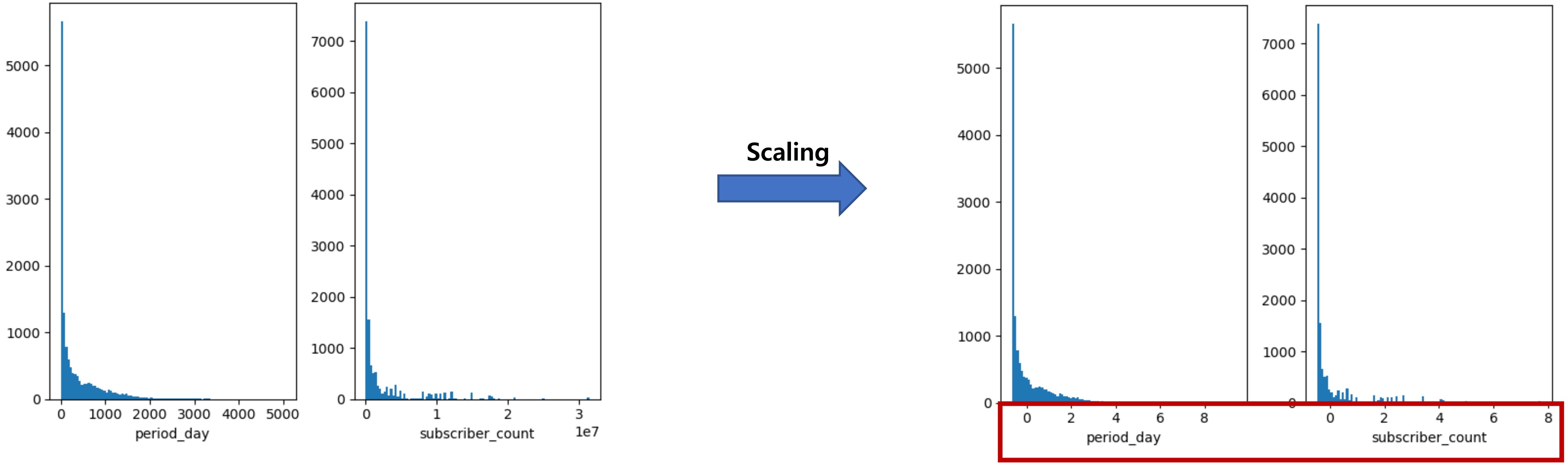
Data z-score Normalization

For **Period** and **Number of Subscribers** data

When learning with small data set without preprocessing, MSE loss popped up in the middle of learning.

We estimated that the cause was that the **range difference** between the two features was very **large**.

Period_day range: 0 ~ 5020 / **Subscriber_count range**: 0 ~ 3.14e+07



Baseline Models

Simple linear regression model
MLP model
Hidden layers: 20492 / 307 / 40 / 1

Experiment Orders

Optimize model architecture
Explore optimal hyperparameters
Train
Evaluate results compared with baseline linear regression and MLP models

05

Evaluation

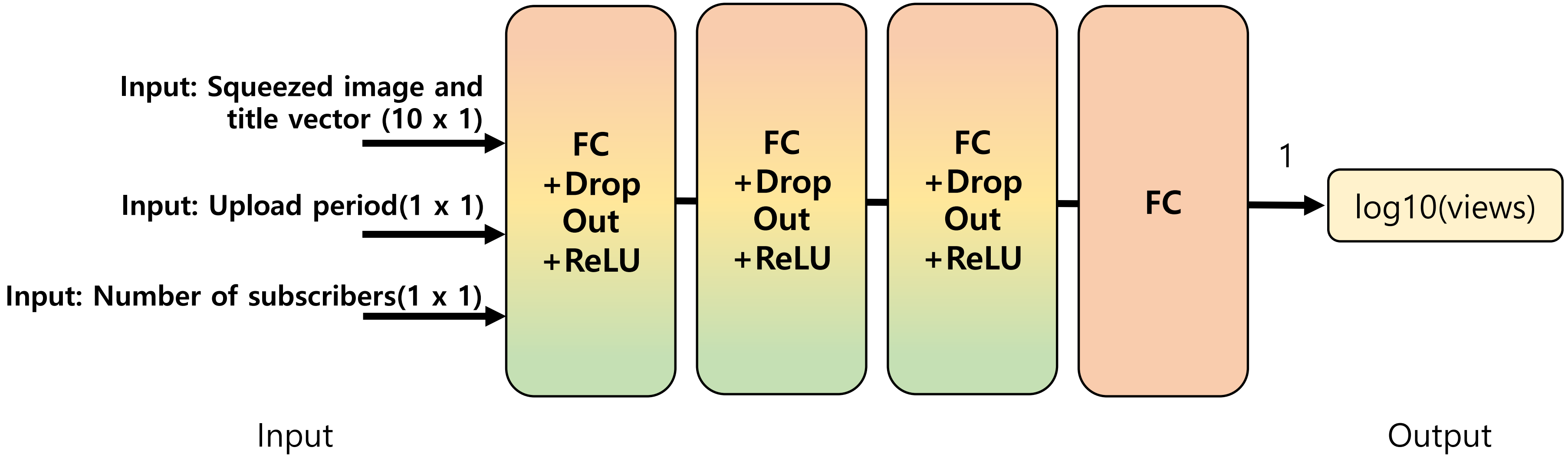
Model Design Configuration

Use small part of dataset: 1600 / 200 / 200 (8 : 1 : 1)

Add DropOut and more Hidden Layers

DropOut rate: 0.3

MSE (Test): 17.21



Hyperparameter Derivation

K Fold Cross Validation, K = 3

- epochs = 20
- optimizer = Adam
- learning rate = 0.001 / **0.0005** / .0001
- weight decay = **0** / 1e-3 / 1e-5

✓ Best MSE(Validation): **4.81**

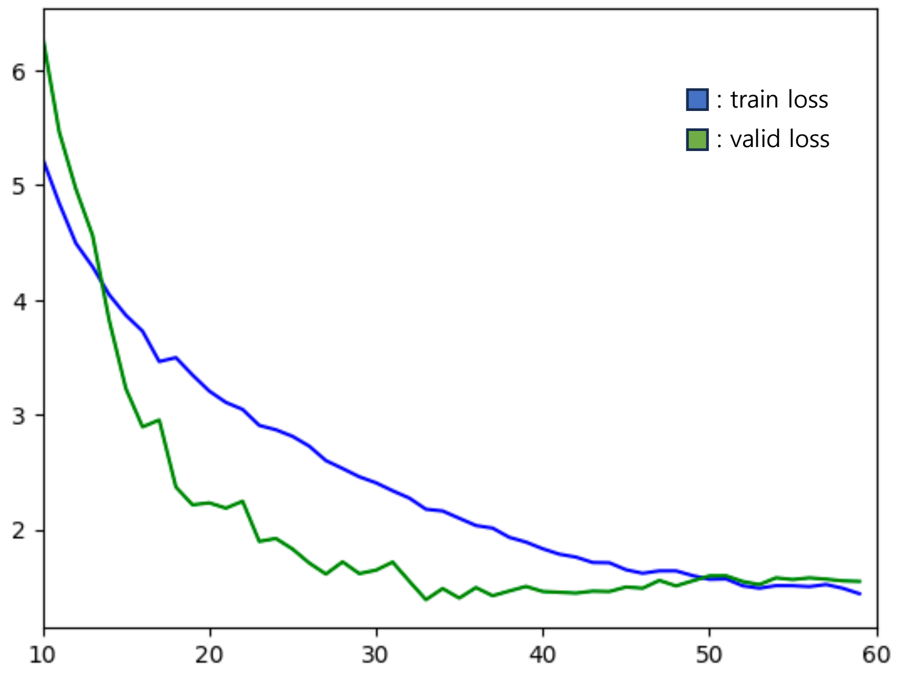
Training Result

Hyper parameters: batch size = 64 / epochs = 60 / learning rate = 0.0005 / weight decay = 0

Early stopping at epoch 50

Small Dataset MSE ($Test_{Small}$) : 17.21

Full Dataset MSE (Test) : **1.32** (▼15.89)

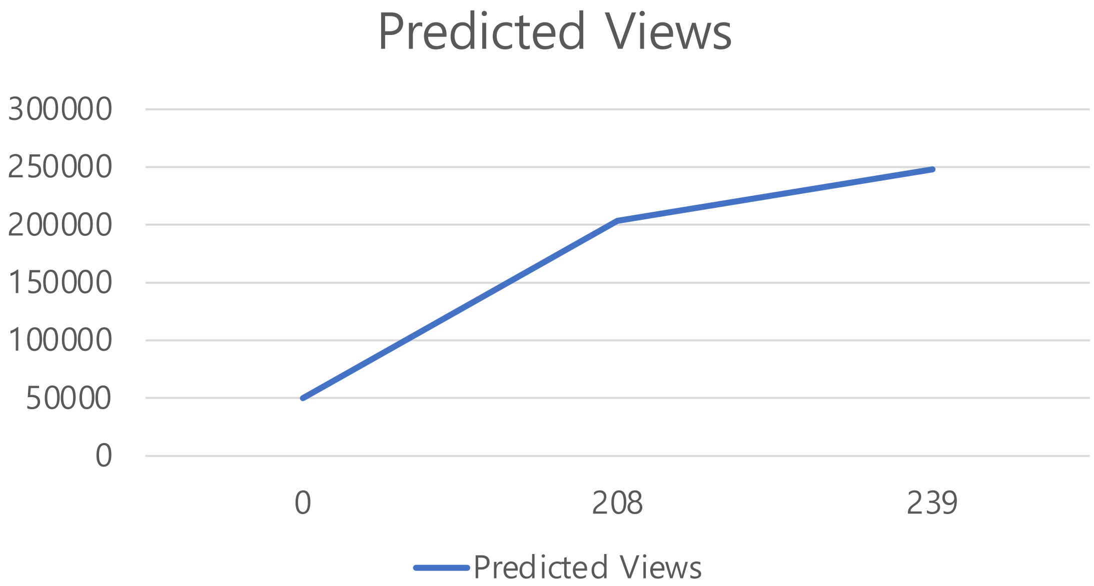


Performance Comparison

	Baseline Linear Regression	Baseline MLP	**Ours**
MSE	2.34	1.63	**1.32**
ΔMSE	▲1.02	▲0.31	**-**
# Parameters	-	6,303,712 (▲8,301)	**6,295,411**

Sample Output

Current View Count: 360K



06

Conclusion

Our novelty

Analysis

The result of the feature importance shows that **utilizing thumbnail feature is important for predicting view counts**

People click upon inherent features of thumbnails

Our model is performes better than baseline MLP model with fewer parameters with MSE of **1.32** (▼1.02 in **logarithm**)

Sophisticated architecture to balance between features