

# 네이버 웹툰 썸네일로 장르 예측하기





# CONTENTS

- 1 주제 선정
- 2 데이터 소개
- 3 방법론
- 4 결론
- 5 한계점 및 의의













# 1 주제 선정



## 웹툰시장규모1조원시대

실사용순위 TOP 10 :

네이버 웹툰, 카카오페이지, 다음웹툰, 라프텔, 레진코믹스, 만화경, 미스터블루, 모두의 웹툰, 투믹스, 케이툰



 네이버 웹툰 - Naver Webtoon	 카카오페이지 KakaoPage
 다음웹툰 - DAUM WEBTOON	 라프텔 - 애니 감상 및 취향저격 추천,...
 레진코믹스 - 매일 무료 웹툰	 만화경 - 온라인 만화책방, 웹툰
 미스터블루 - 최신 웹툰 소설 만화 무...	 모두의 웹툰
 투믹스 - 웹툰 (무료웹툰/인기만화)	 케이툰(KTOON) - 무료 웹툰







# 1 주제 선정

## 다양한 장르

에피소드    오니버스    스토리    |    일상    개그    판타지    액션    드라마    순정    감성    스릴러    무협/사극    스포츠

에피소드 웹툰 총 192

인기순    업데이트순    조회순    별점순     

 <p>대학원 탈출일지 요다 2022.05.19 ★★★★★ 9.98 전체보기</p>	 <p>독립일기 자끼 2022.05.21 ★★★★★ 9.97 전체보기</p>	 <p>비즈니스 여친 루즌아/서해 2022.05.17 ★★★★★ 9.80 전체보기</p>
 <p>쇼미더럭키짱! 박태준/김성모 2022.05.19 ★★★★★ 9.44 전체보기</p>	 <p>한림체육관 혜성/이석재 2022.05.16 ★★★★★ 9.74 전체보기</p>	 <p>약한영웅 서패스/김진석 2022.05.21 ★★★★★ 9.83 전체보기</p>

작품의 장르 - 작가가 가장 잘 파악

장르에 대한 구분의 인식이 그리 높지 않거나, 작가도 선택하기 애매모호한 경우 존재

웹툰 콘텐츠를 중심에 둔 글로벌 시장, IP 활용, 독자들의 선호도 측면 활용도 높음

# 1 주제 선정

웹툰 썸네일을 여러 타입으로 만들어 A/B 테스트 진행

-> 최대 클릭률과 최소 클릭률의 차이 8.38%p



웹툰 썸네일은 매체와 장르의 특성을 반영한 고유의 디자인 패턴을 형성

# 1 주제 선정

썸네일로 장르 예측하기!!!

총 10개 장르

일상 개그 판타지 액션 드라마 순정 감성 스릴러 무협/사극 스포츠



자극적인

액션  
스릴러  
무협/사극

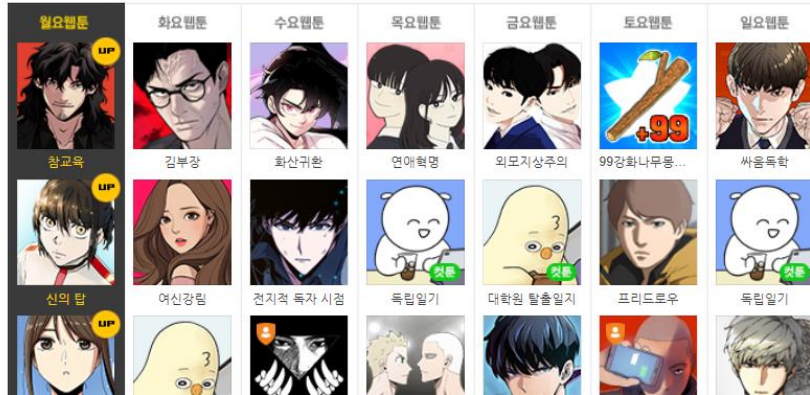
자극적이지 않은

드라마  
순정  
감성  
일상  
개그  
판타지  
스포츠

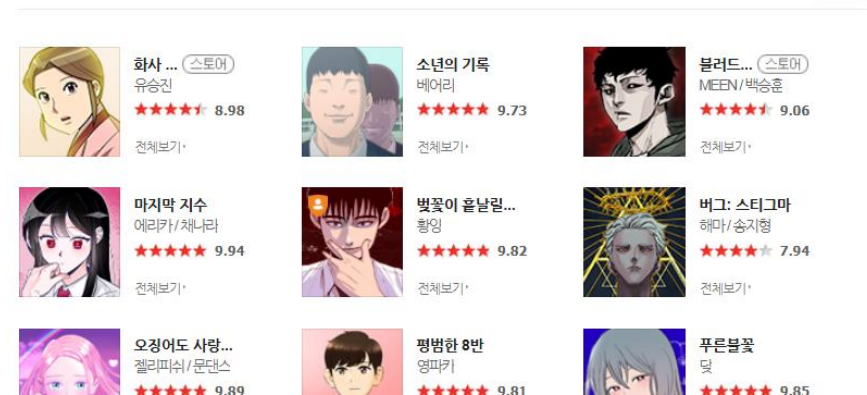
## 2 데이터 소개

### 데이터 수집

요일별 전체 웹툰



완결 웹툰



△ 연재중인 요일별 전체 웹툰

△ 완결 웹툰



**참교육** 채용택 / 한가람

무너진 교권을 지키기 위해 교권보호국 소속 나화진의 참교육이 시작된다!  
<부활남> 채용택 작가 X <신석기녀> 한가람 작가의 신작!

스토리, 액션 | 15세 이용가

⊕ 관심웹툰

첫회보기

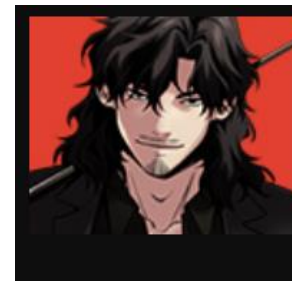
작가의 다른 작품 ▾

❤ 19,952

## 2 데이터 소개

### 데이터 수집

	title	author	genre	thumbnail
0	참교육	채용택 / 한가람	스토리, 액션	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
1	쇼미더럭키짱!	박태준 / 김성모	에피소드, 액션	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
2	뷰티풀 군바리	설이 / 윤성원	스토리, 드라마	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
3	신의 탑	SIU	스토리, 판타지	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
4	퀘스트지상주의	박태준 만화회사	스토리, 드라마	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
...	...	...	...	...
1981	AA	선정, 지속	스토리, 로맨스	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
1982	사랑in	전세훈	옴니버스, 로맨스	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
1983	까뱃	X-TEAM	에피소드, 개그	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
1984	와랑카	우주인	에피소드, 개그	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>
1985	타투	노지현	옴니버스, 드라마	<a href="https://shared-comic.pstatic.net/thumb/webtoon...">https://shared-comic.pstatic.net/thumb/webtoon...</a>



### 데이터 라벨링 (0 : 자극적인, 1 : 자극적이지 않은)



0



1

Train : 469

Valid : 59

Test : 59

Train : 444

Valid : 56

Test : 57



## 2 데이터 소개

### ■ 전처리

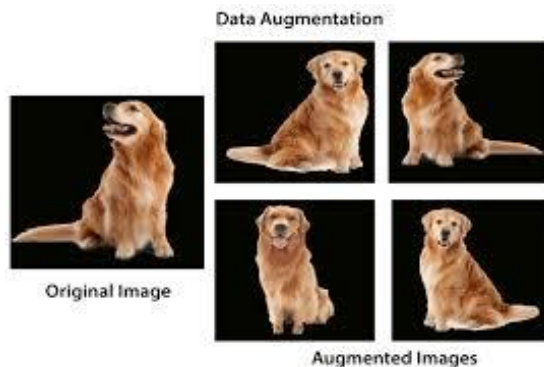
#### [Input data set 조정]

- target\_size : (150, 150)
- Class\_mode : binary
- Batch\_size : 20

#### [이미지 좌표값 scaling : 0~1]



#### [data augmentation]



- rotation\_range : 40
- Width\_shift\_range : 0.2
- Height\_shift\_range : 0.2
- Shear\_range : 0.2
- zoom\_range : 0.2
- Horizontal\_flip : True
- Fill\_mode : nearest

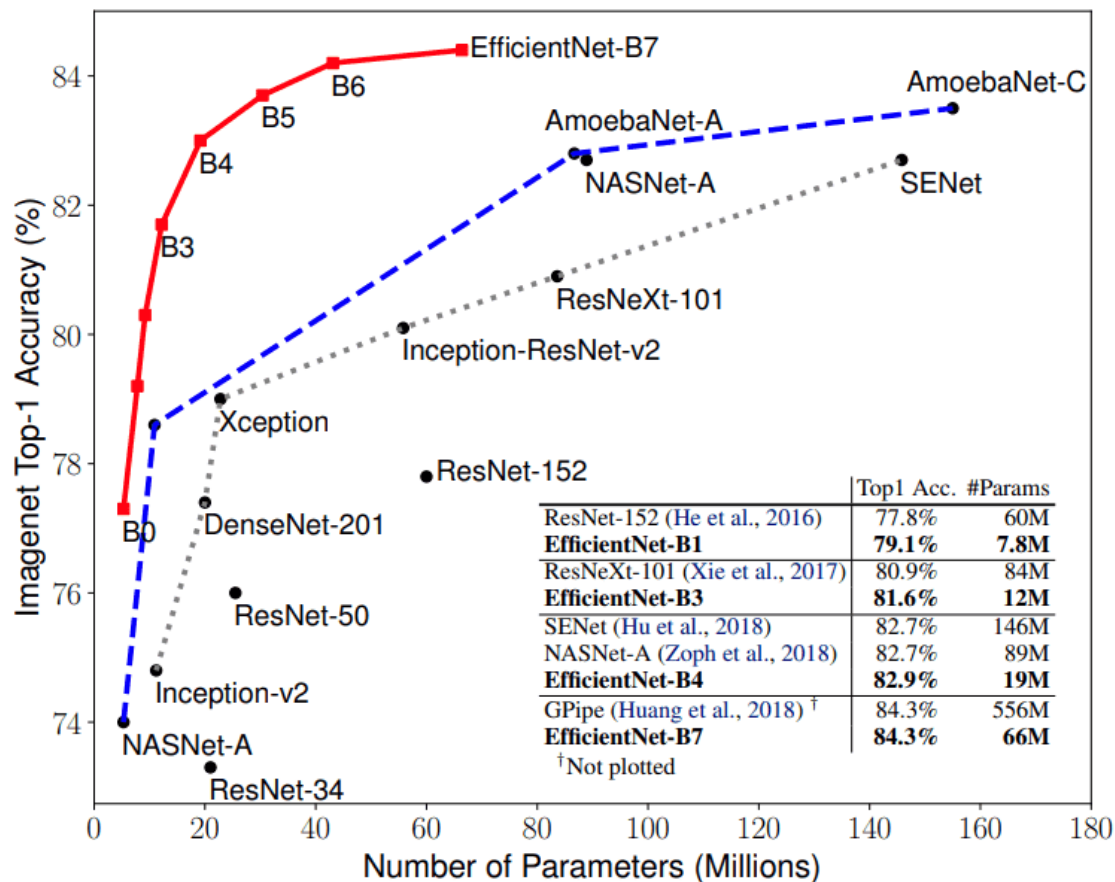
Train - Found 913 images belonging to 2 classes.

Valid - Found 115 images belonging to 2 classes.

Test - Found 116 images belonging to 2 classes.

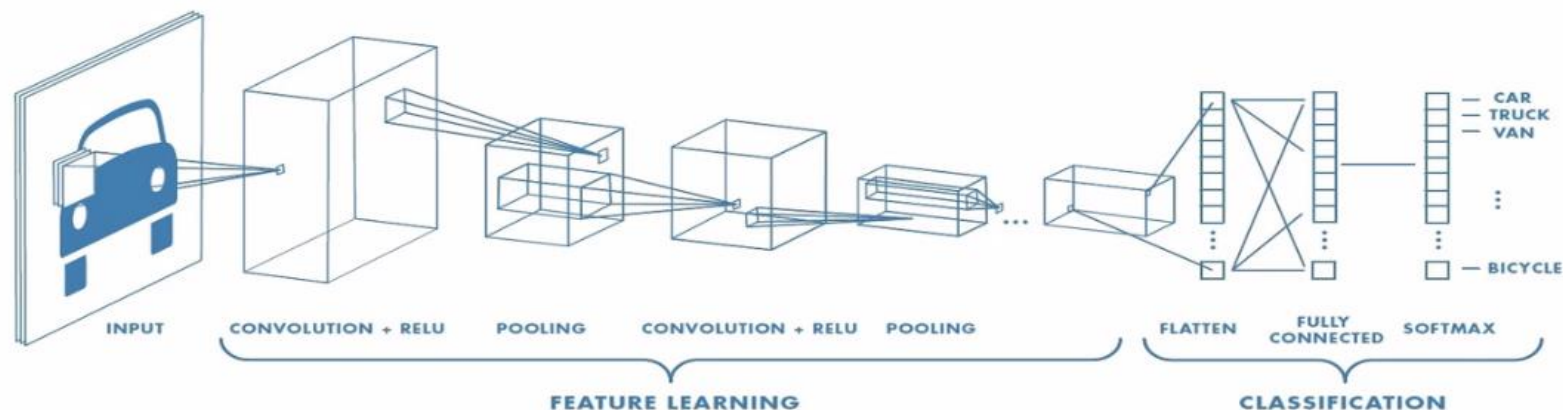
### 3 방법론

CNN, VGG, ResNet, EfficientNet



### 3 방법론

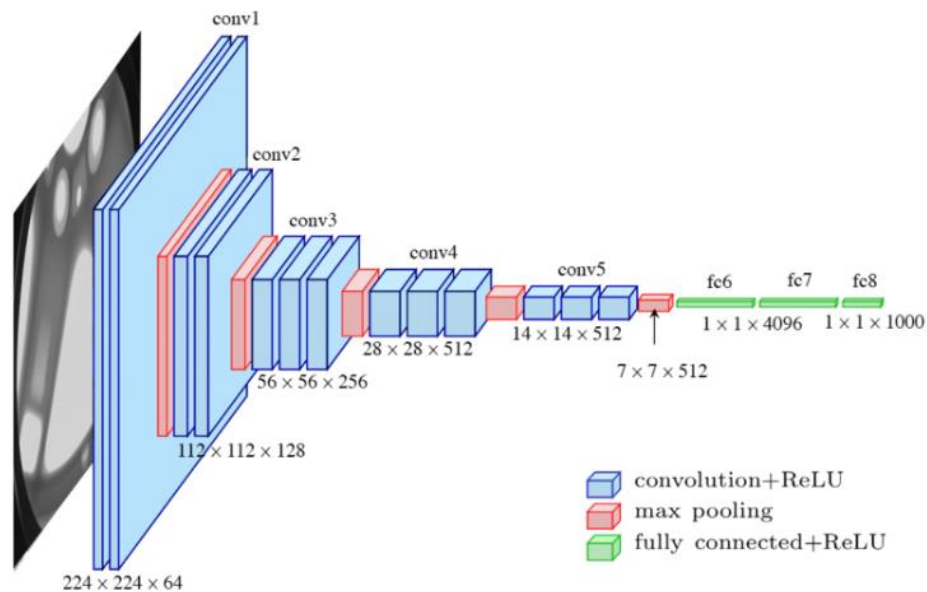
#### CNN



- Convolutional Neural Network, 합성곱 신경망
- 인간의 시신경 구조를 모방한 기술
- 특징맵 생성하는 필터까지 학습 가능해 vision분야에서 성능 우수
- 이미지 인식 패턴 찾는 데 유용
- 이미지의 공간 정보 유지한 채 학습
- 사람이 여러 데이터 보고 기억한 후 맞추는 것과 유사

# 3 방법론

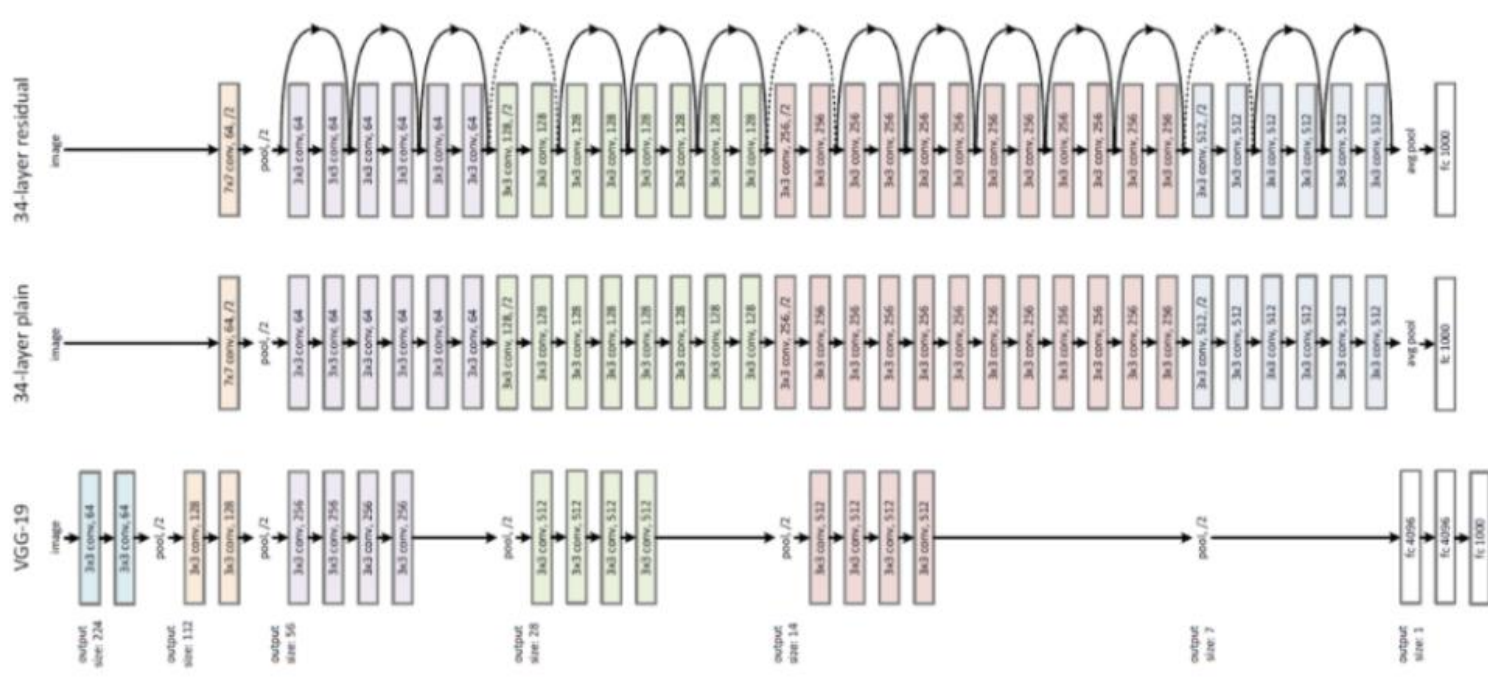
## VGGNet



- 작은 필터를 사용함으로써 더 많은 ReLU함수를 사용할 수 있고 더 많은 비선형성 확보 가능
- VGGNet은  $3 \times 3$ 의 작은 필터를 모든 Conv 레이어에 사용
- VGGNet의 A~E까지 각각의 다른 모델이 아니라 업그레이드된 모델

# 3 방법론

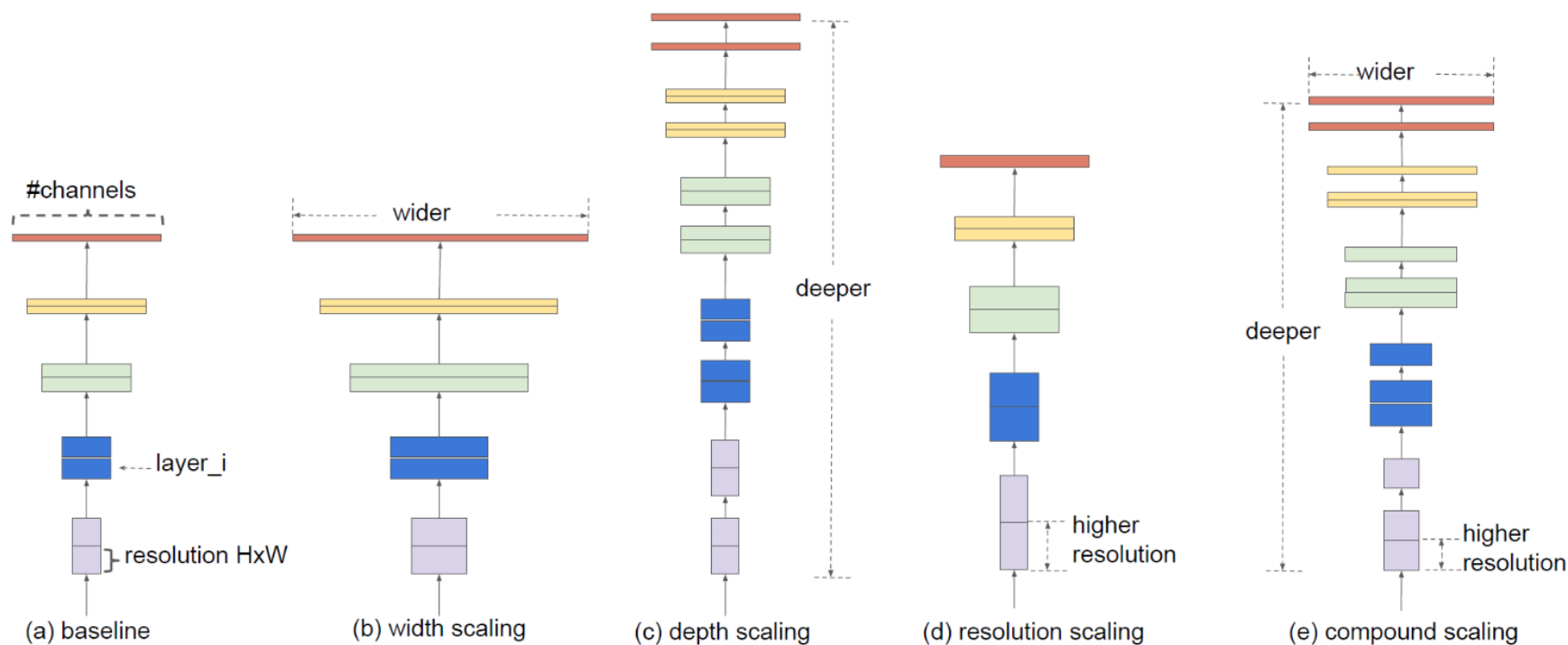
## ResNet



VGGNet의 8배인 152 layer를 자랑하며, 여기에 앙상블 기법을 적용해 오차를 3.75%까지 줄임

### 3 방법론

#### EfficientNetB0



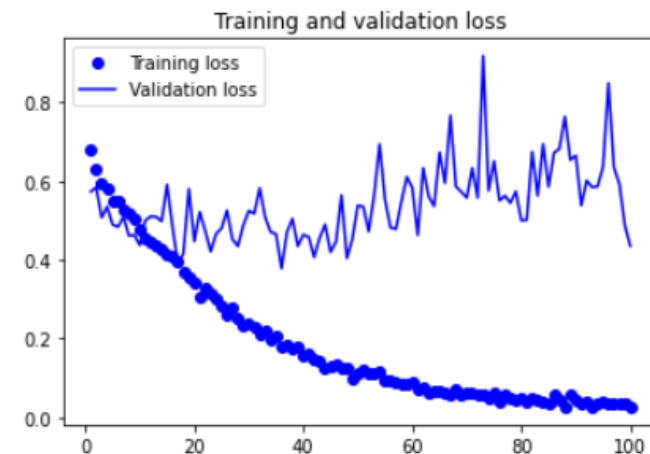
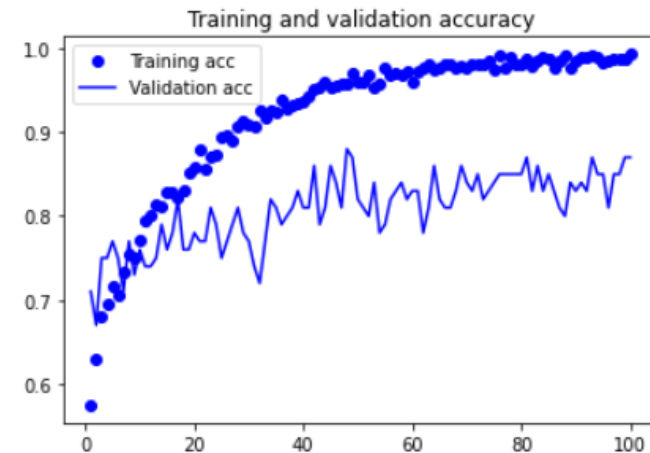
최적의 조합을 AutoML을 통해 찾은 논문. 조합을 효율적으로 만들 수 있도록 하는 compound scaling 방법을 제안하며 이를 통해 더 작은 크기의 모델로도 SOTA를 달성한 논문

# 4 결론

## ■ 모델 파라미터

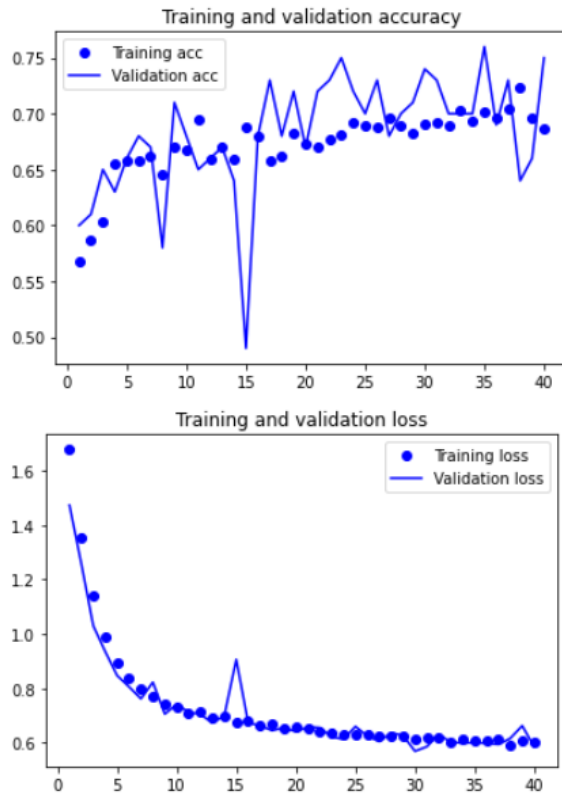
- Loss : binary\_crossentropy
- Metrics : accuracy
- Steps per epoch : 45
- Validation\_steps : 5
- Epochs : 50

## ■ VGGNet+finetuning epoch 100-> 과적합



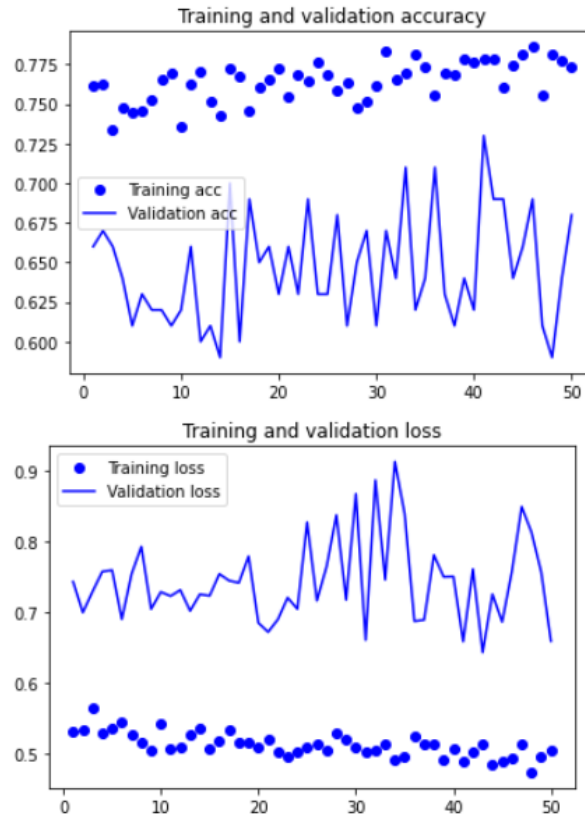
# 4 결론

## CNN



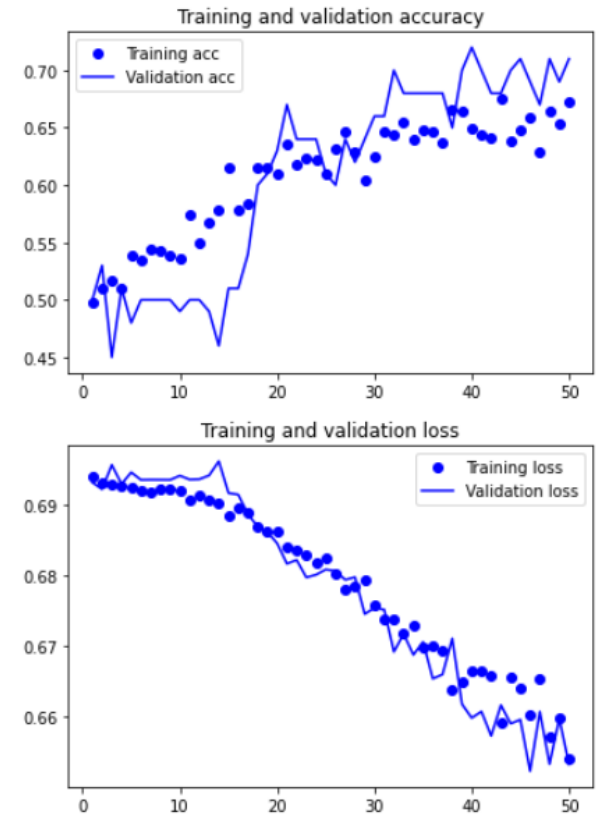
Optimizer : RMSprop(lr= $1e-4$ )

## ResNet



Optimizer : RMSprop(lr= $2e-5$ )

## EfficientNet



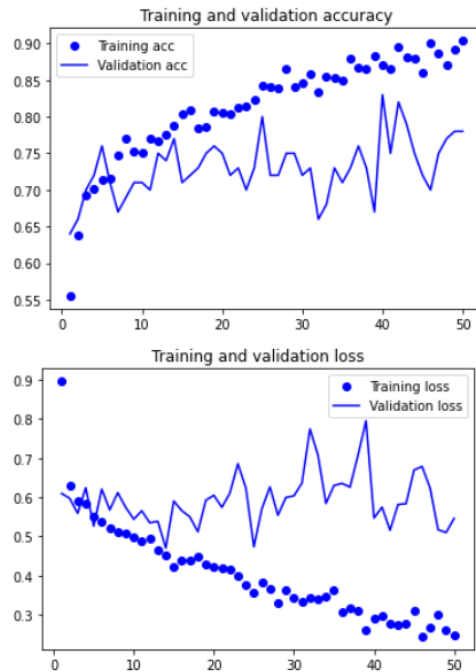
Optimizer : Adam(lr= $1e-5$ )



# 4 결론

## VGGNet

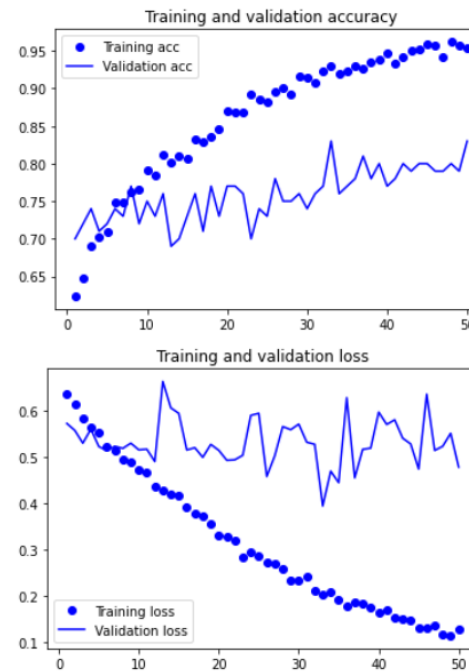
Imagenet으로 사전 훈련된 VGGNet에서 Fully connected layer만 직접 쌓은 모델



Optimizer : Adam(lr=0.001)

## VGGNet+finetuning

Imagenet으로 사전 훈련된 VGGNet에서 Fully connected layer만 직접 쌓고 사전 훈련된 모델의 마지막 층인 'block5\_conv1'의 가중치 학습 가능



Optimizer : RMSprop(lr=1e-4)

## 4 결론

### ■ 모델 비교하기

	Train_loss	Train_acc	Valid_loss	Valid_acc	Test_loss	Test_acc
CNN	0.6024	0.6865	0.5852	0.7500	0.5831	0.7414
VGG16	0.2477	0.9037	0.5460	0.7800	0.5739	0.76521
VGG16 + finetuning	<b>0.1288</b>	<b>0.9541</b>	<b>0.4781</b>	<b>0.8300</b>	<b>0.4447</b>	0.8261
<b>ResNet</b>	0.5048	0.7637	0.7646	0.6100	0.7218	<b>0.6348</b>
EfficientNet B0	0.6542	0.6719	0.6533	0.7100	0.6531	0.7043

## 5 한계점 및 의의

### 한계점

1. 현저히 적은 데이터셋  
-> 추후 다양한 다른 플랫폼의 웹툰을 연계하여 다양한 장르로 진행 가능
2. 장르의 군집화 필요 (비슷한 장르 많음, 데이터 불균형 문제)

### 의의

1. 썸네일 뿐만 아니라 줄거리도 고려한 장르 예측 가능
2. 이를 이용한 사용자 기반 맞춤형 웹툰 추천 시스템에 적용 가능
3. 어린이용 웹툰 필터링 가능

감사합니다.

