VGGNet 의 이해

Very Deep Convolutional Networks for Large-Scale Image Recognition

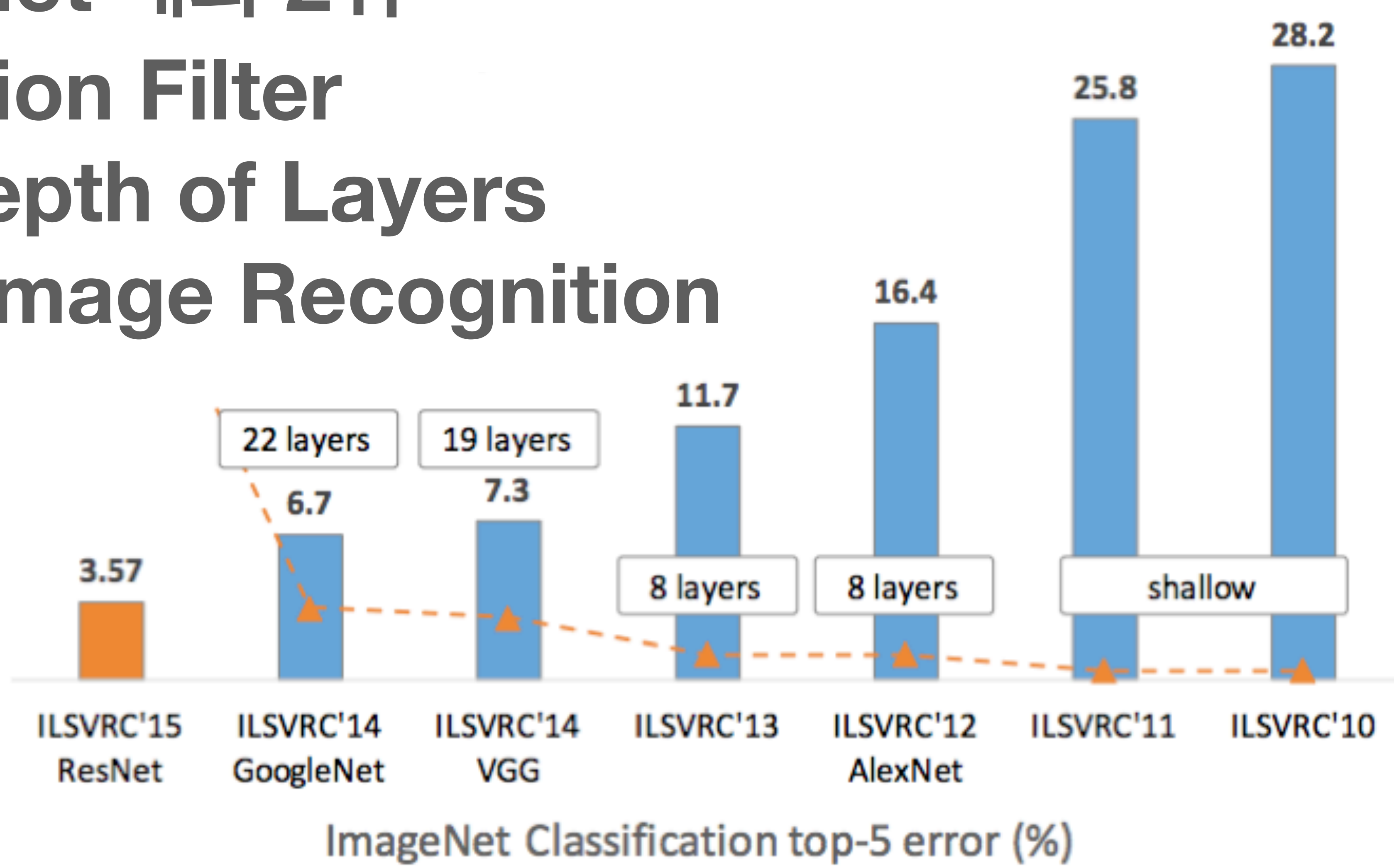
혁신성장 - 인공지능 큐레이터 양성과정 - 홍승우

Agenda

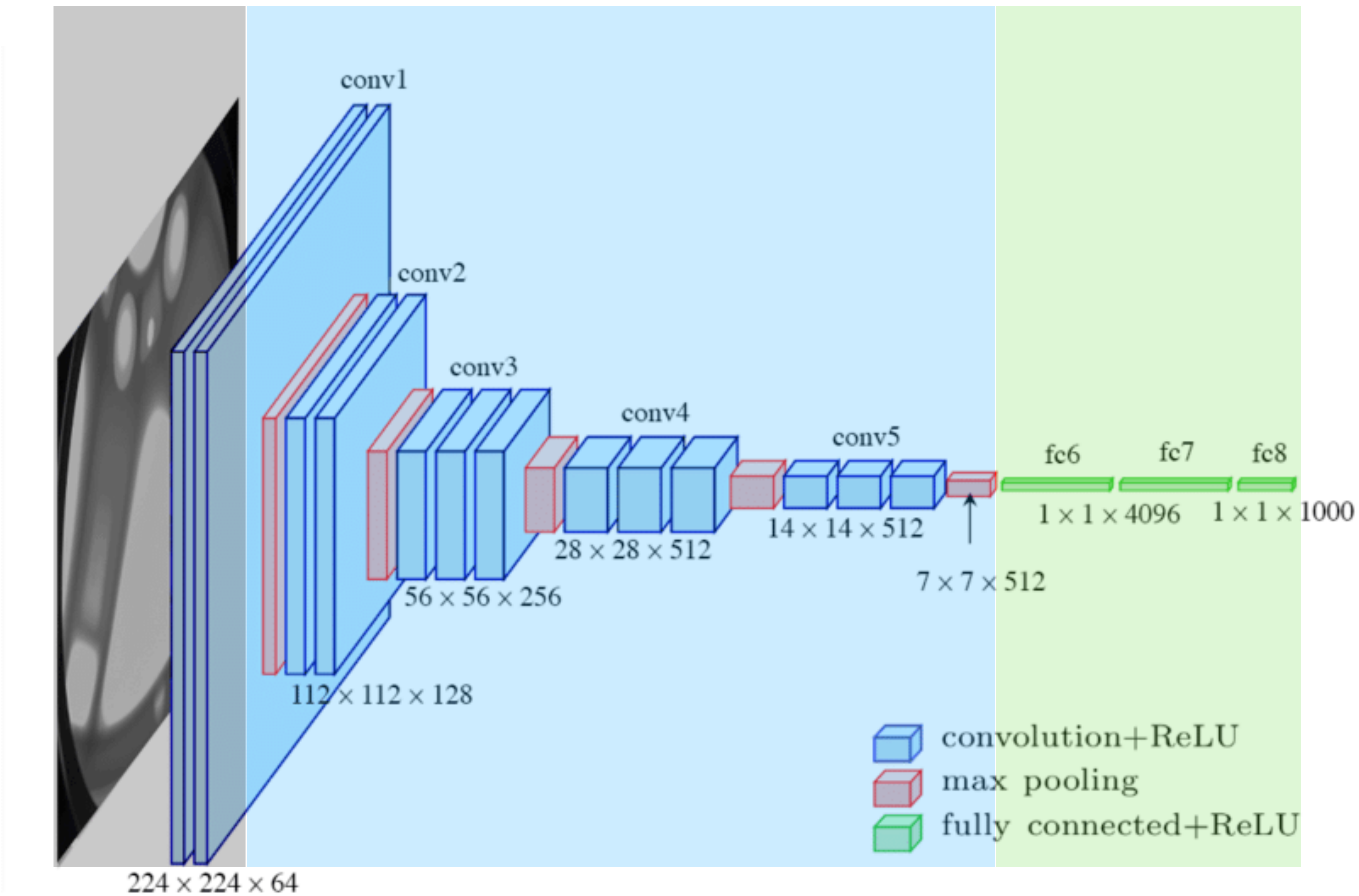
- 개요
- 기본 구조 (Architecture)
- Kernel의 특징
- 네트워크의 깊이
- Training & Testing
- Results
- 결언

개요

- Simonyan & Zisserman, 2014
- 2014 Image Net 대회 2위
- 3x3 Convolution Filter
- Increasing Depth of Layers
- Large Scale Image Recognition



Architecture (VGG-16)



224x224x3
이미지 입력

Convolution & Max Pooling 반복

- **Filter 크기: 3x3 고정**
- **Padding = 'same' 이미지 사이즈 변화 없음**

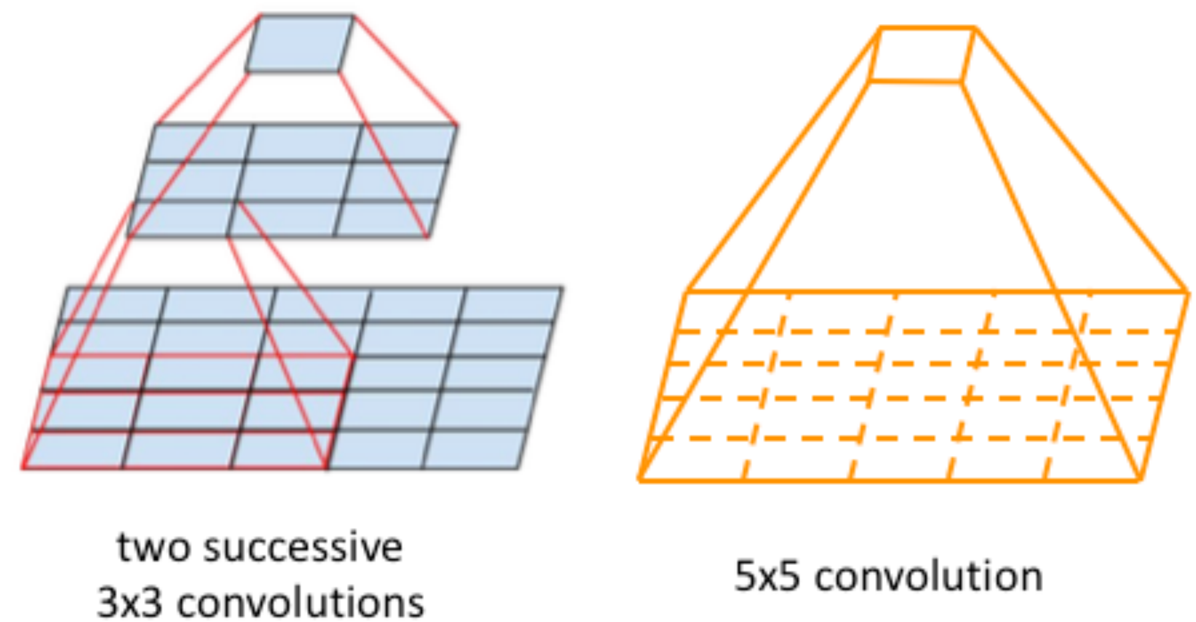
FC Layer

Conv 3x3 Filter vs. 5x5 Filter

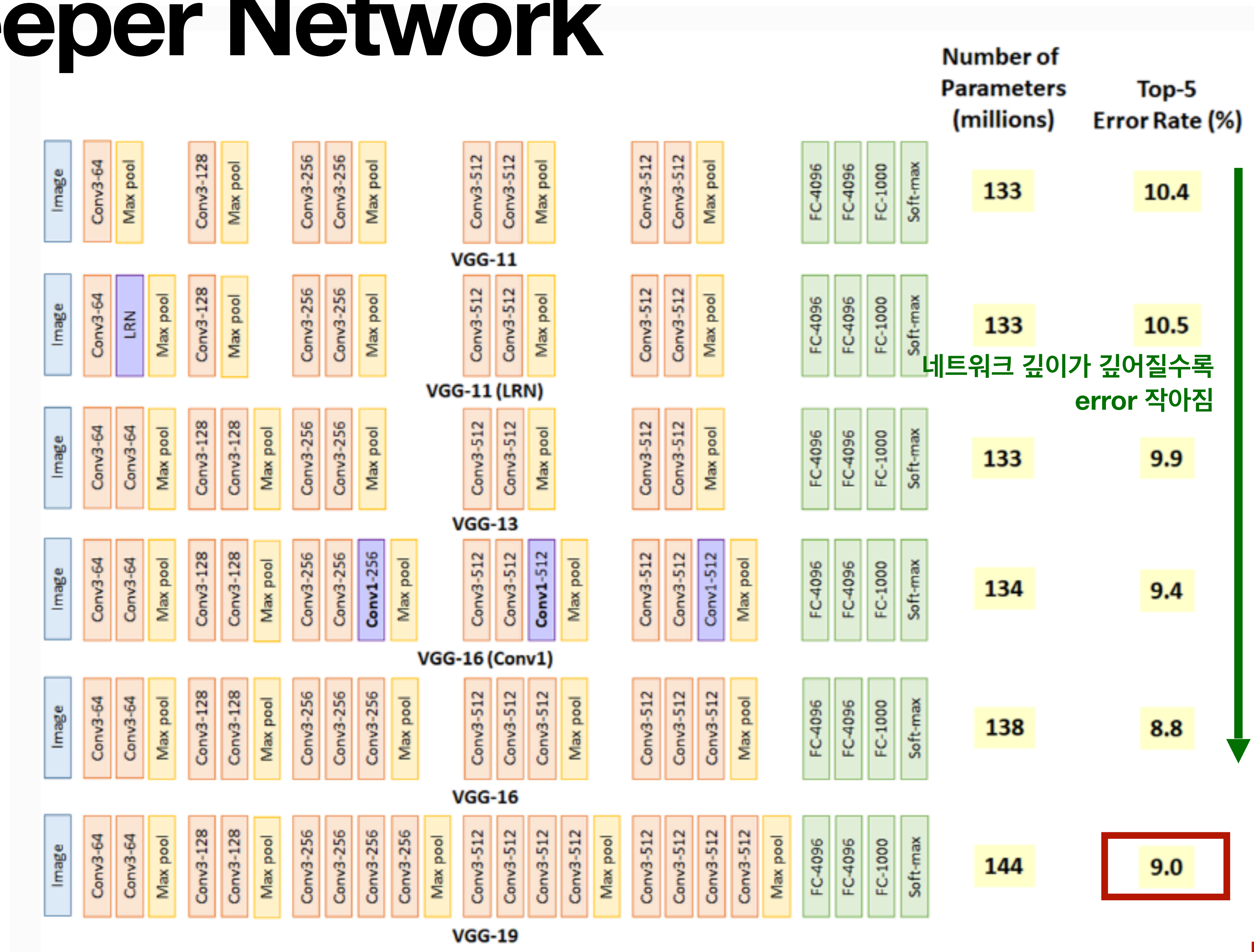
- 동일한 effective receptive field

- **파라미터 수 감소**
  → 빠른 학습속도
  → Overfitting 감소

- **더 깊은 network**
  → 더 많은 ReLu 의 사용
  → Non-linearity 증가
    → Better feature selection



Deeper Network



Error 값이
비슷하거나
더 나빠짐

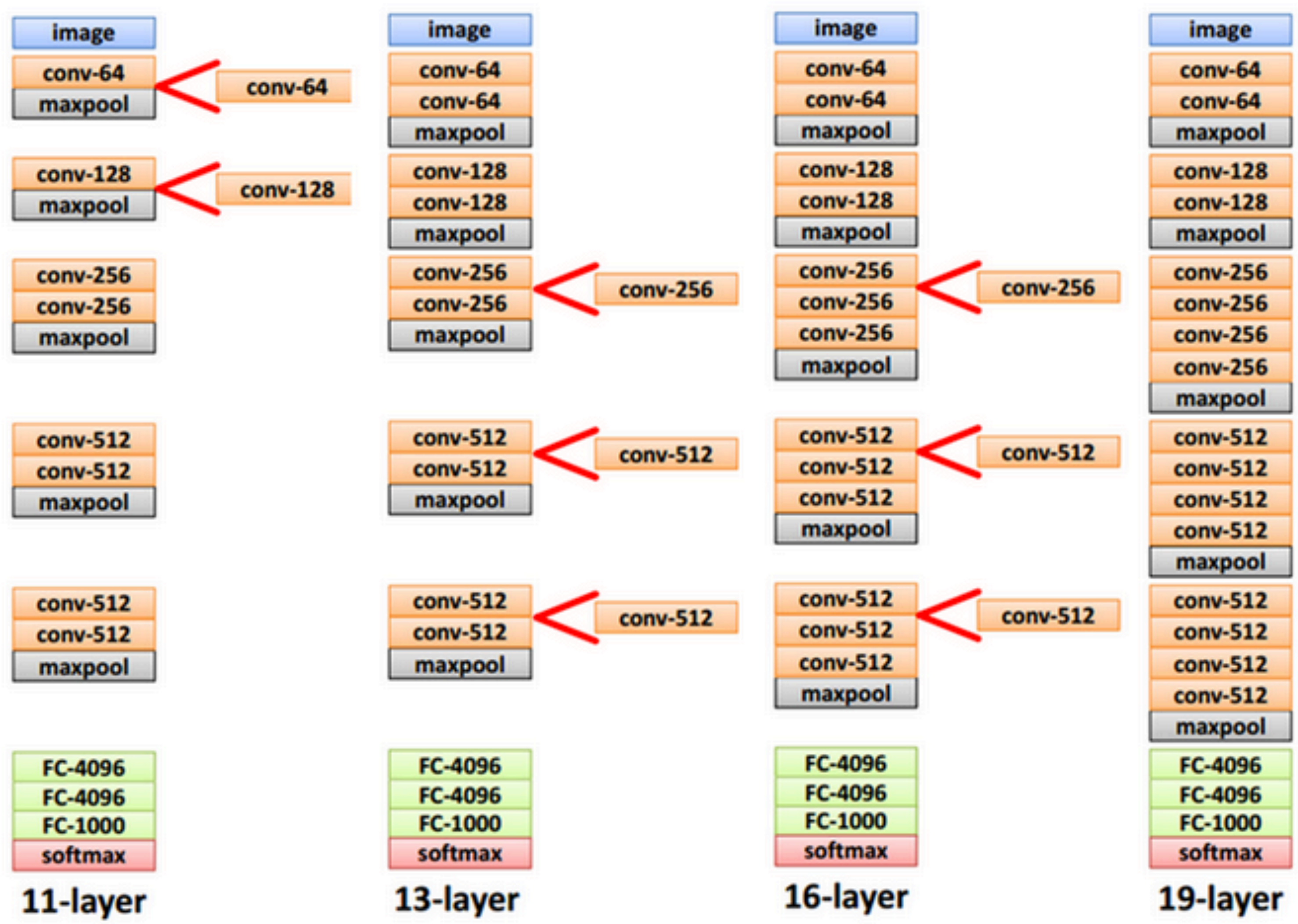
Train & Test

- **Data augmentation**
  - ILSVRC - 1000 Classes, Class 당 약 1000장
  - Scaling → 좌우반전 & RGB 조작

- **Training**
  - Pre-initialization
  - Single Scale → Multi Scale (Scale Jittering)

- **Testing**
  - Single / multi test scale
  - Multi-crop / Dense / Fusion

Training - Pre-initialization



- 11 layer 모델 먼저 학습
- 기존 학습된 파라미터를 기반으로, 레이어 추가하여 학습

- 학습 시간 감소
- Vanishing gradient 문제 해결

Training - Parameters

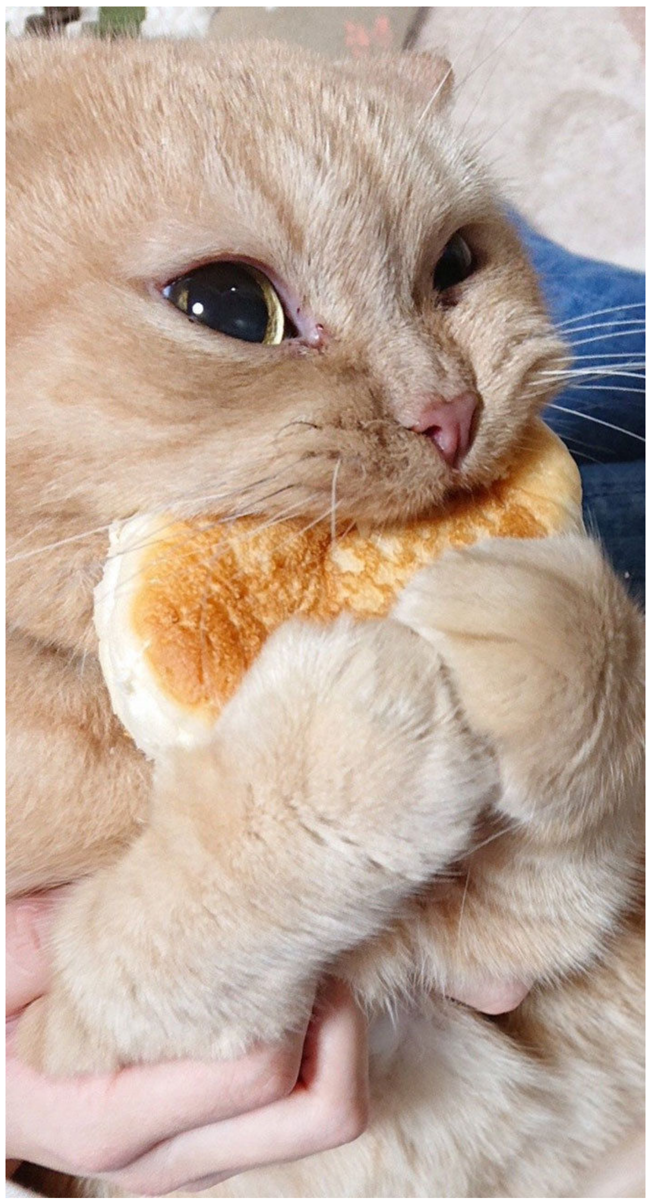
- **Hyper Parameters**
  - Cost function: Multinomial logistic regression
  - Batch size: 256
  - Optimizer: Momentum = 0.9
  - Regularization: L2 regularization = $5 \cdot e{-4}$
  - Dropout = 0.5
  - Learning rate: e-2

- **Trainable Parameters**

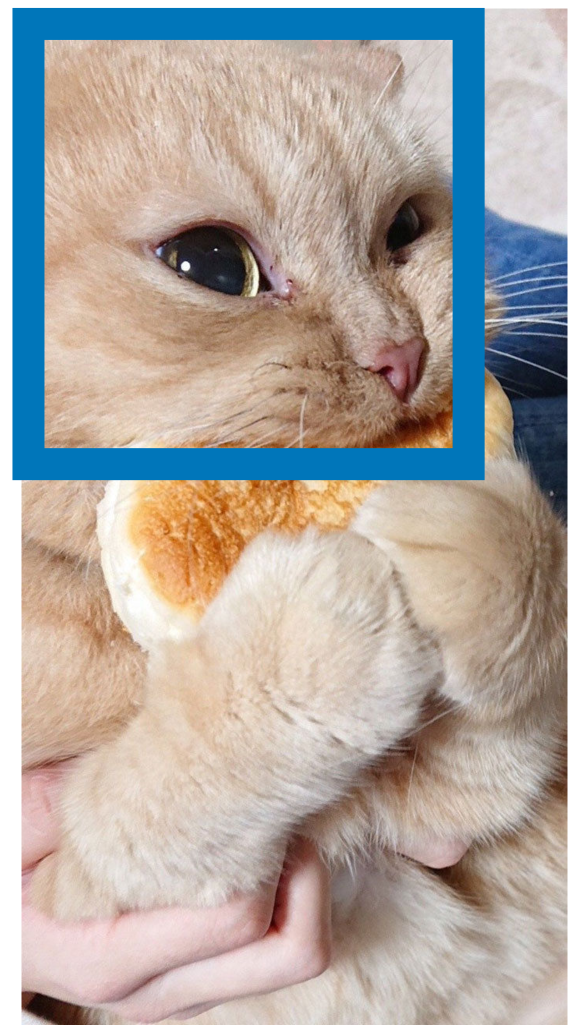
Network	A, A-LRN	B	C	D	E
# params	133M	133M	134M	138M	144M

Single Scale

Isotropically-Rescaled

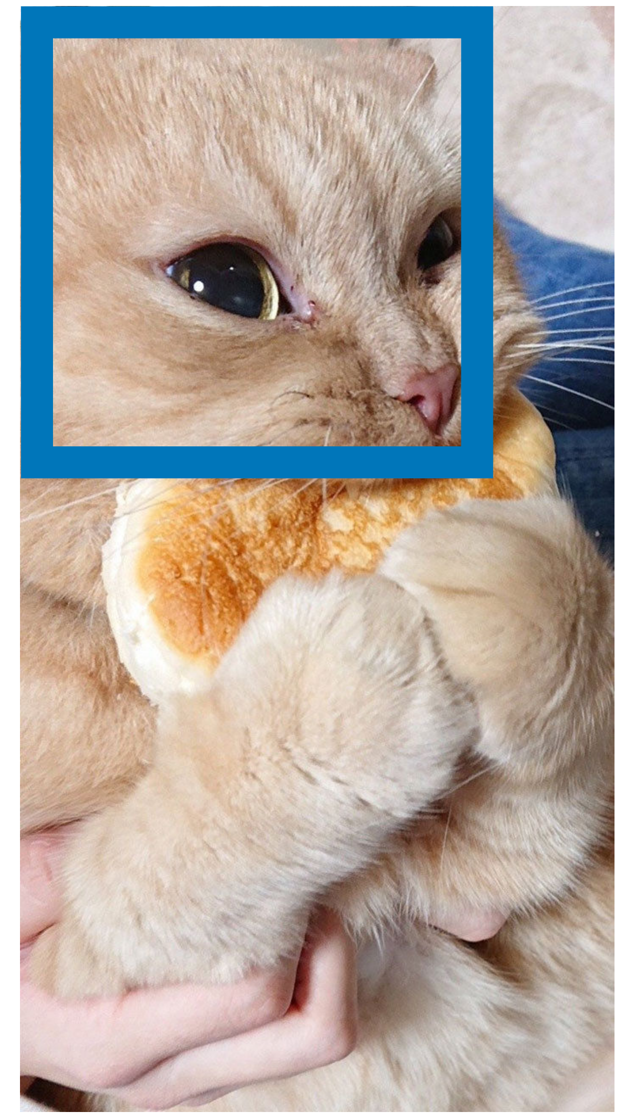


Original Image
512x930



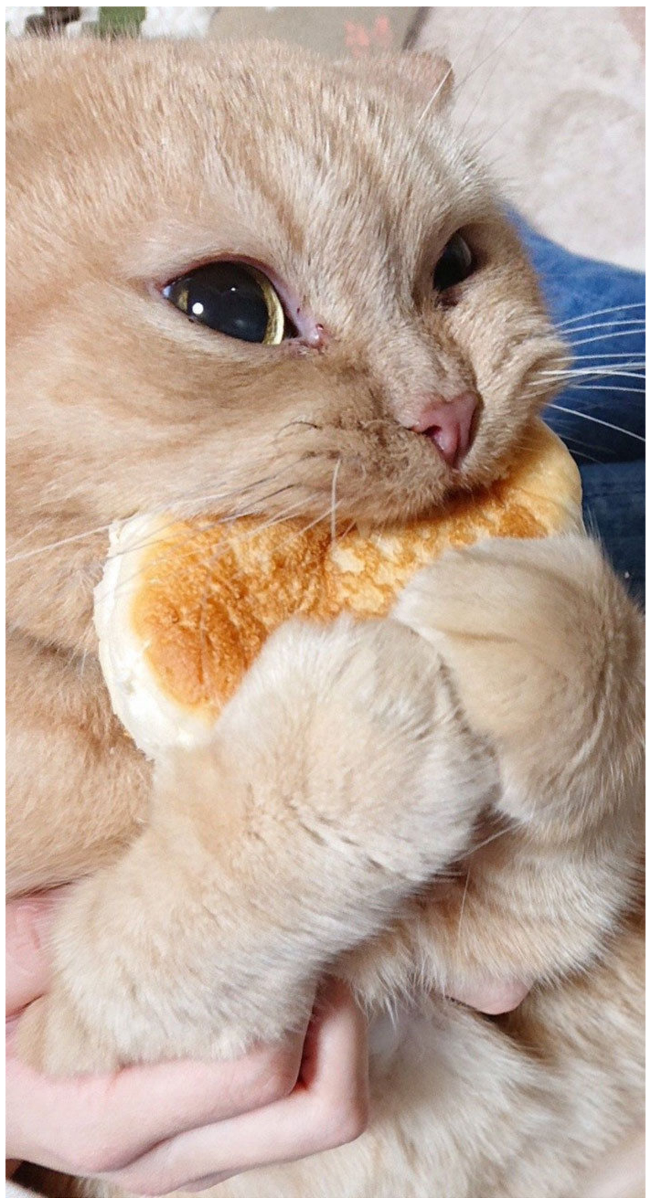
S=256
256x465

or



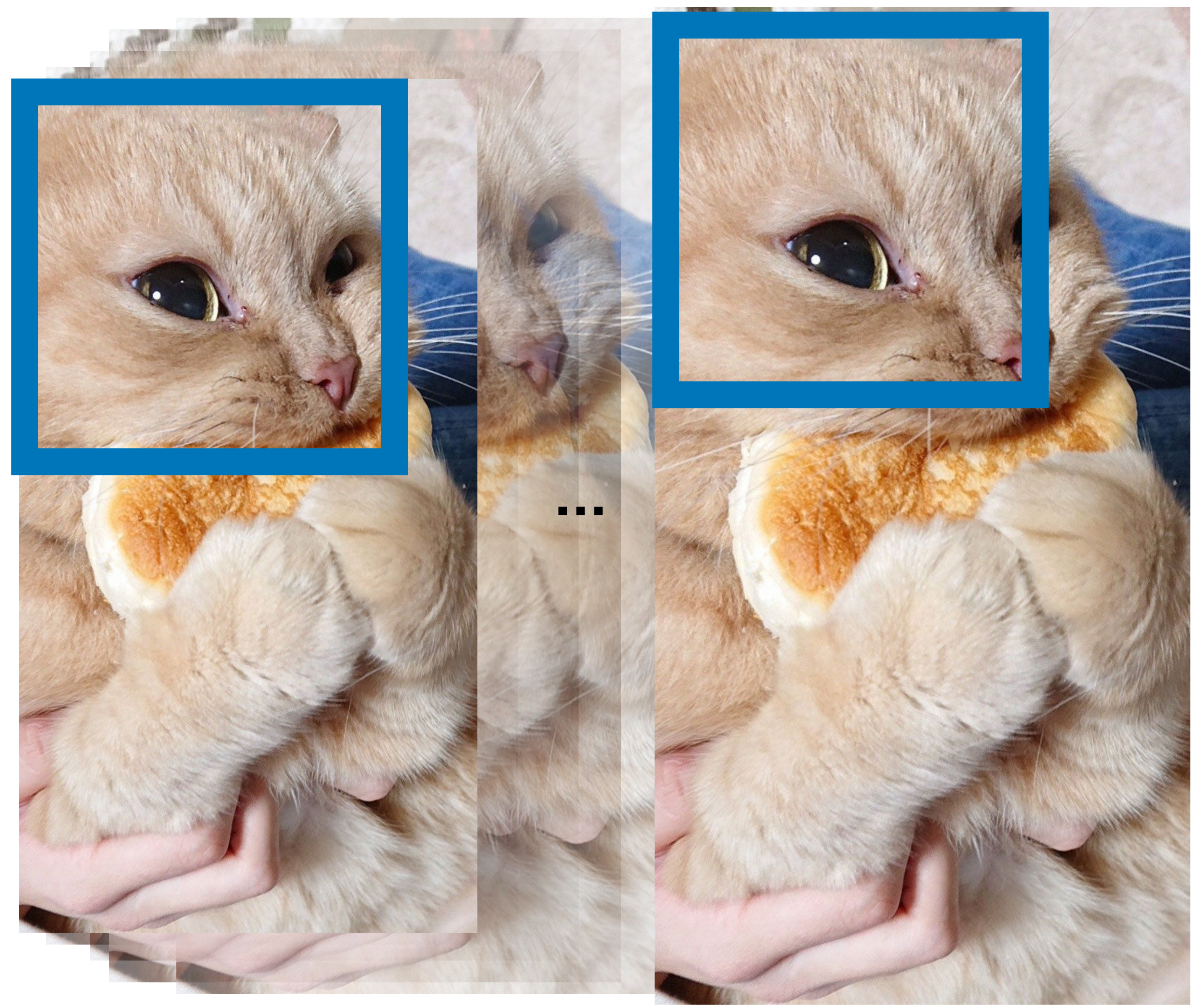
S=384
384x698

Multi Scale



Original Image
512x930

Isotropically-Rescaled



S=256
256x465

...

S=512
512x930

Multi-crop & Dense Evaluation

- **Multi-crop evaluation**
  - 50 crops / (5x5, 2flips) x 3 scales = **Total 150 crops**
  - Voting
  - 이미지수 증가에 따른 **연산량 증가**

- **Dense evaluation**
  - Max Pooling 을 밀집하게 (densely) 적용하여 **resolution의 떨어짐을 보완**
  - 연산량 측면에서 효율적
  - 픽셀 간격의 크기 문제로 인한 학습 결과 떨어짐

- **ConvNet Fusion**
  - **Multi-crop + Dense**
  - 동시에 사용함으로써 각각의 단점 보완

Testing - FC layer ⇢ Conv. layer

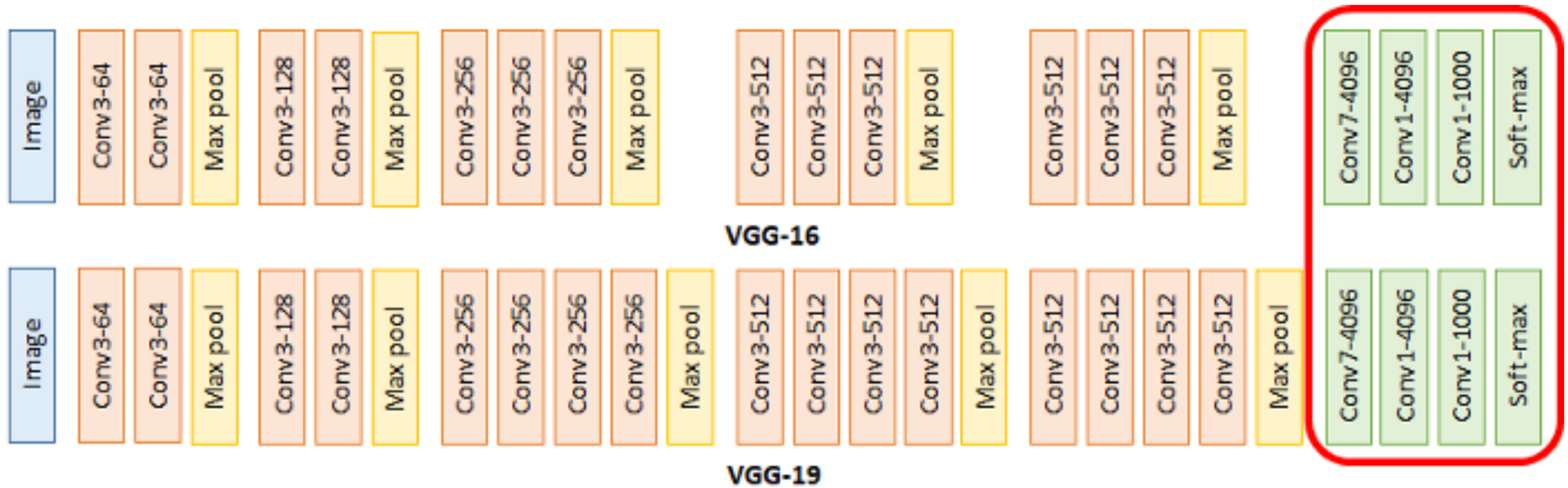
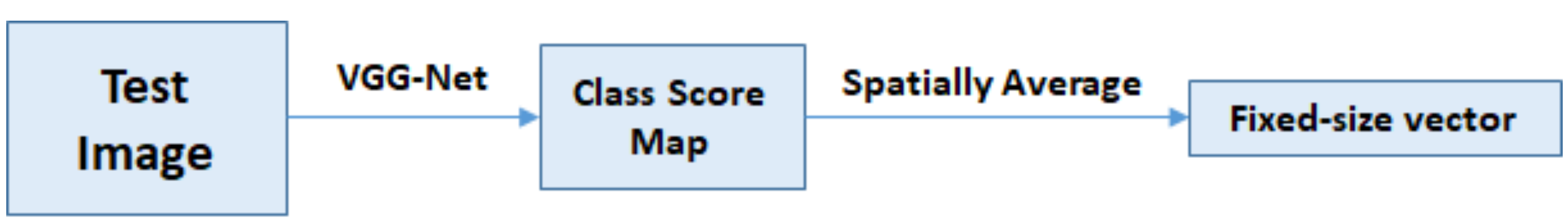


Image scale 과 상관없이 class 를 구분할 수 있도록, FC layer를 Conv. layer로 변경



최종 feature 가 1x1 이 아니더라도, 평균을 내어 1x1 feature map 생성

Test 결과 비교

Table 3: **ConvNet performance at a single test scale.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
A	256	256	29.6	10.4
A-LRN	256	256	29.7	10.5
B	256	256	28.7	9.9
C	256	256	28.1	9.4
	384	384	28.1	9.3
	[256;512]	384	27.3	8.8
D	256	256	27.0	8.8
	384	384	26.8	8.7
	[256;512]	384	25.6	8.1
E	256	256	27.3	9.0
	384	384	26.9	8.7
	[256;512]	384	**25.5**	**8.0**

Table 4: **ConvNet performance at multiple test scales.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
B	256	224,256,288	28.2	9.6
C	256	224,256,288	27.7	9.2
	384	352,384,416	27.8	9.2
	[256; 512]	256,384,512	26.3	8.2
D	256	224,256,288	26.6	8.6
	384	352,384,416	26.5	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**
E	256	224,256,288	26.9	8.7
	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**

Single test scale

- Test 이미지 사이즈 고정
- S=Single Scale,
  - S=Q
- S=Multi Scale,
  - 0.5x(256+512) = 384 고정

Multi test scale

- 하나의 S 사이즈에 대해 여러 test image 사용
- S=Single Scale,
  - Q={S−32,S,S+32}
- S=Multi Scale,
  - Q={$Smin$,0.5($Smin$+$Smax$),$Smax$}

Test 결과 비교

Table 5: **ConvNet evaluation techniques comparison.** In all experiments the training scale S was sampled from $[256; 512]$, and three test scales Q were considered: $\{256, 384, 512\}$.

ConvNet config. (Table 1)	Evaluation method	top-1 val. error (%)	top-5 val. error (%)
D	dense	24.8	7.5
	multi-crop	24.6	7.5
	multi-crop & dense	**24.4**	**7.2**
E	dense	24.8	7.5
	multi-crop	24.6	7.4
	multi-crop & dense	**24.4**	**7.1**

Dense, Multi-crop
& ConvNet Fusion

- Image Size 세팅은 모두 동일
  - S=Multi Scale
  - Q={256, 384, 512}

Multi Scaling Training, multi scaling testing,
dense 와 multi-crop 을 모두 사용할 때,
가장 좋은 결과를 보여줌

결언

- Network depth의 영향력을 확인하기 위해
  3x3 Convolution을 활용한 단순한 구조의 모델

- 다양한 경우에 대한 수치를 발표하며
  Deep CNN에 대한 이해를 도움

- 파라미터의 수가 매우 많아
  학습 시간이 오래걸리는 문제점

감사합니다.

References

1. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint *arXiv:1409.1556.*
2. *VGGNet — Organize everything I know documentation.* (n.d.). Read the Docs. Retrieved September 23, 2020, from https://oi.readthedocs.io/en/latest/computer_vision/cnn/vggnet.html
3. Tsang, S. (2020, September 1). *Review: VGGNet — 1st Runner-Up (Image Classification), Winner (Localization) in ILSVRC 2014.* Medium. https://medium.com/coinmonks/paper-review-of-vggnet-1st-runner-up-of-ilsvlc-2014-image-classification-d02355543a11
4. Do-Woo-Ner, D. (2020, January 23). *7. VGGNet.* Time Traveler. https://89douner.tistory.com/61?category=873854