

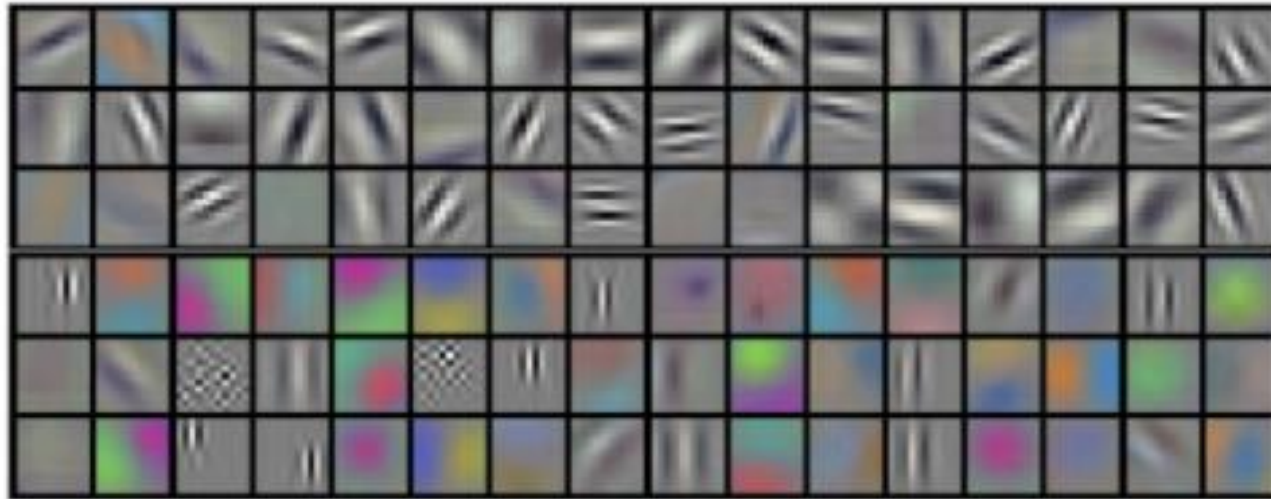
# Aggregated Residual Transformations for Deep Neural Networks

# Background

- 대규모 데이터에서 신경망을 통해 학습된 기능은 훈련 중에 **사람의 개입을 최소화** 해야 하며 **다양한 인식 작업으로 이식될 수** 있습니다.
- VGG와 ResNet은 **같은 모양의 블록을 여러 개 쌓는** 것으로 하이퍼 파라미터를 제한하여 단순화 하였다. Inception은 성능은 좋지만 채널이 여러 개로 나뉘 각기 다른 연산을 거치기에 모델의 구조는 복잡하다.
- Inception의 구조처럼 **채널이 여러 개로 나뉘 연산**을 하며, VGG와 ResNet처럼 **단순한 구조**를 가지는 모델을 제안했다.

# Motivation

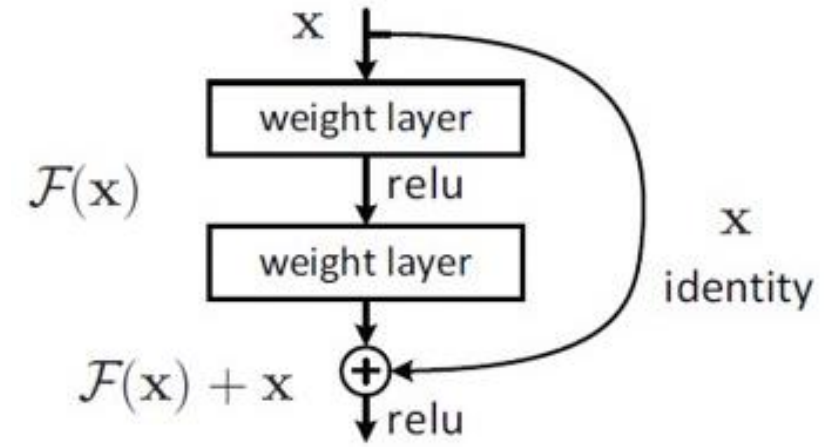
- VGGnet에서 GPU의 한계로 인해 신경망의 **채널 수를 두 개로 쪼개서** 2개의 GPU로 연산했다. 이렇게 연산을 했더니 아래 그림과 같이 서로 다른 특징(흑백, 컬러)에 집중하여 학습이 되었다. 그리고 이를 합쳐서 테스트를 진행하였을 때 모델의 성능이 향상됐다. 이를 Grouped convolution 이라고 한다.



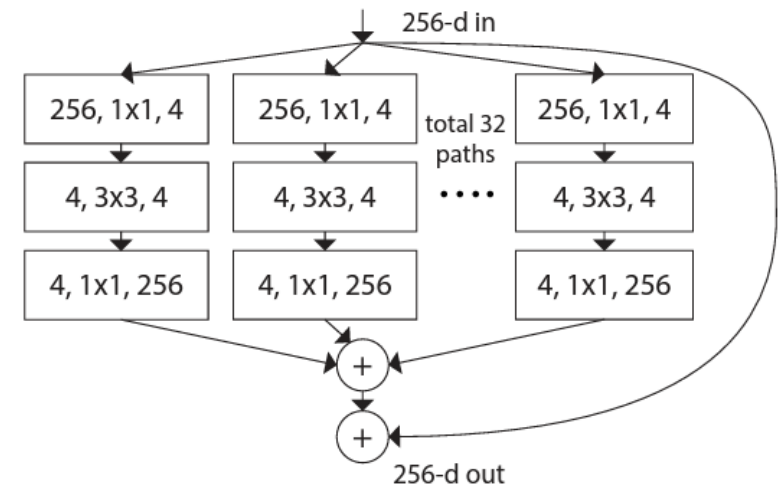
<나뉜진 신경망의 피쳐맵 시각화>

# Model

- ResNet은 Bottleneck에서 convolution 연산을 하는 것에 반해 ResNext는 Inception에서 사용하던 Cardinality 개념을 도입하여 **채널을 N개만큼 쪼개서 모두 같은 convolution 연산 후 Weight를 더해**주는 Split-Transform-Merge 구조를 사용한다.
- 이 때, 쪼개진 CNN이 몇개의 path를 가지는지를 결정하는 하이퍼 파라미터가 바로 Cardinality이며, 각각의 path에서 가지는 채널을 depth라고 정의한다.



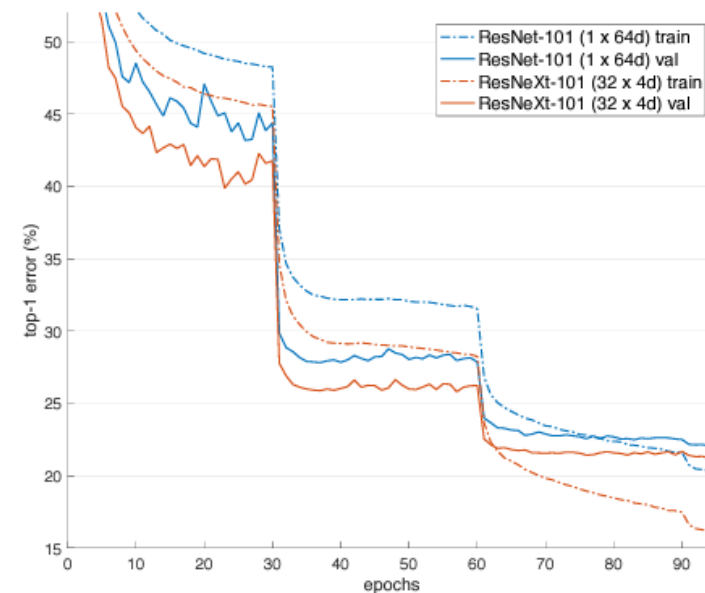
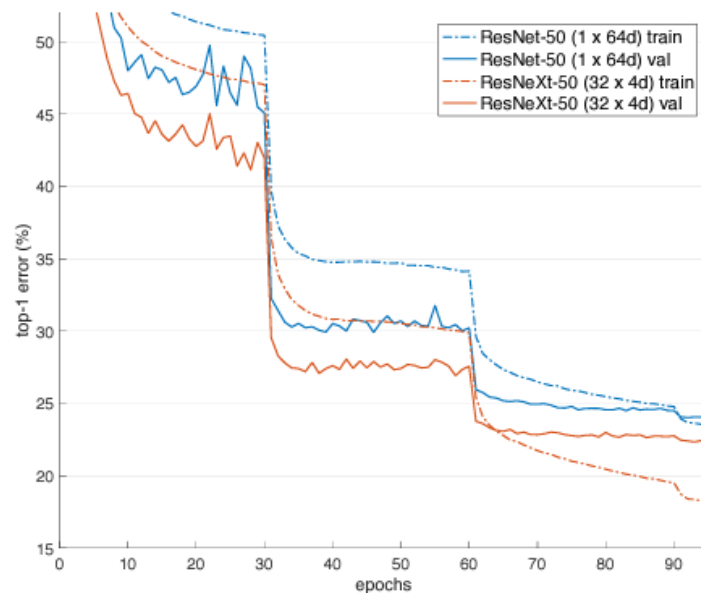
<Residual function 구조>



< Split-Transform-Merge 구조 >

# Result

- ResNet과 같은 깊이를 가지는 ResNeXt와 ResNet을 비교하였을 때 비슷한 파라미터, 연산량을 가지면서 더 좋은 성능을 보였다.



<같은 깊이를 가진 ResNet과 ResNeXt 학습, 검증 비교>

	ResNet-50	ResNeXt-50 (32×4d)
# params.	25.5×10 <sup>6</sup>	25.0×10 <sup>6</sup>
FLOPs	4.1×10 <sup>9</sup>	4.2×10 <sup>9</sup>

<ResNet과 ResNeXt 파라미터, 연산 속도 비교>