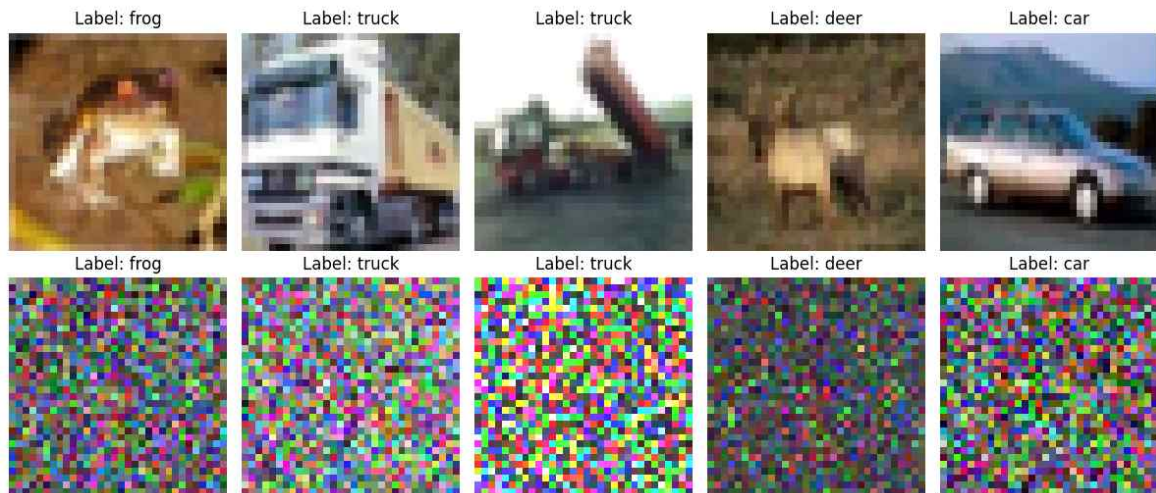


단순 FC로만 구성된 모델과 CNN이 이미지를 어떻게 받아들이는지 비교

Cifar10은 RGB 3채널 데이터셋이기 때문에 RGB별로 랜덤한 좌표로 픽셀을 이동시켜서 학습에 적용시켜보기(모든 이미지에 대해서 동일한 랜덤 좌표로 이동시킴.)

Original vs Randomized Pixel CIFAR-10 Samples



위 이미지가 원본, 아래가 랜덤으로 섞은 이미지: 원본이미지가 어두우면 shuffled 이미지도 어두운 것을 확인할 수 있음. 인간이 보았을 때 shuffled 이미지는 형상을 도무지 알아볼 수 없음. 따라서 실험전 공간적 구조를 학습하지 않는 FC 모델의 경우는 학습에 성공할 수 있을 것이라고 판단했고, CNN은 공간적 구조를 학습하지 못하고 노이즈만 학습하기 때문에 학습 실패할 것이라고 판단함.

## 랜덤 셔플 학습 FC

[Original] 테스트 데이터셋 Loss: 2.7288, 정확도: 13.95%

[Randomized] 테스트 데이터셋 Loss: 1.3526, 정확도: 52.71%

랜덤하게 섞인 이미지를 학습한 FC 모델은 랜덤하게 섞인 이미지도 어느정도 학습하여 구분할 수 있음을 확인. 원본 이미지를 통과시켰을 때는 정확도가 13.95%로 단순 찍었을 때의 확률과 비슷함을 확인할 수 있음.

Predictions by Model Trained on Randomized Data



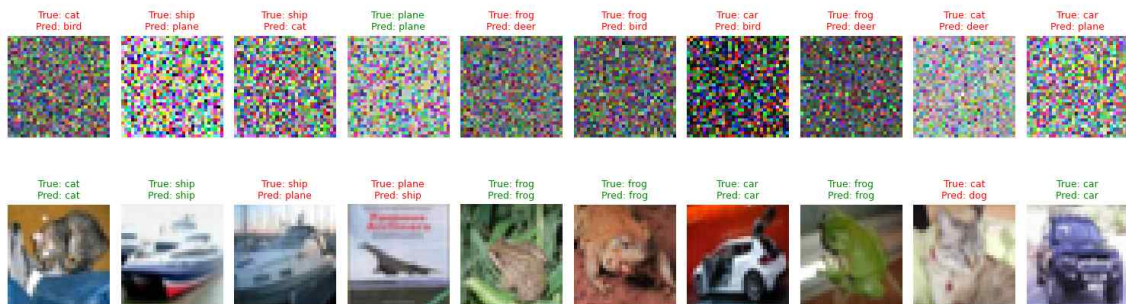
## 원본 이미지 학습 FC

[Original] 테스트 데이터셋 Loss: 1.3352, 정확도: 53.68%

[Randomized] 테스트 데이터셋 Loss: 2.8640, 정확도: 11.42%

원본 이미지를 학습한 FC 모델의 경우에서도 셔플된 이미지를 학습한 모델과 성능이 비슷함을 확인할 수 있음. shffled 이미지를 통과시켰을 때 정확도가 11.42%로 짝은 것과 비슷함.  
(근소하게 원본이미지 학습한 FC가 더 좋음, 52.71% < 53.68%, shffled < original)

Predictions by Model Trained on Original Data



## 랜덤 셔플 학습 CNN

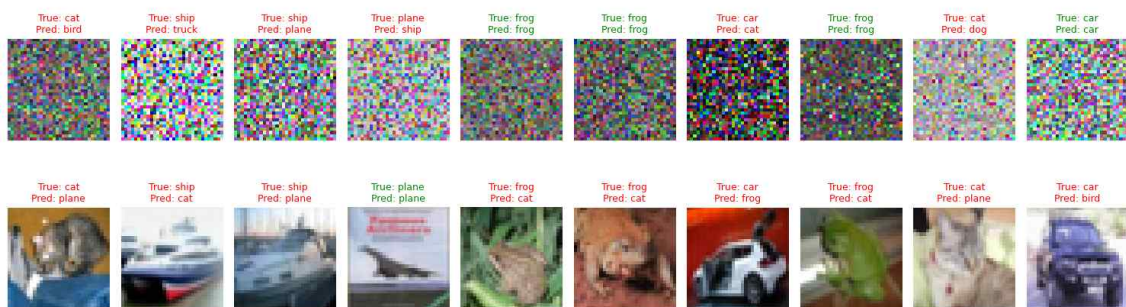
[Original] 테스트 데이터셋 Loss: 5.9565, 정확도: 10.95%

[Randomized] 테스트 데이터셋 Loss: 2.2718, 정확도: 45.31%

예상과 다르게 셔플된 이미지도 CNN은 학습할 수 있었음. 하지만 FC 모델과 비교하면 성능이 더 떨어졌음.

(셔플 학습한 FC가 더 좋음, 52.71% > 45.31%, shffled FC > shffled CNN)

Predictions by Model Trained on Randomized Data



## 원본 이미지 학습 CNN

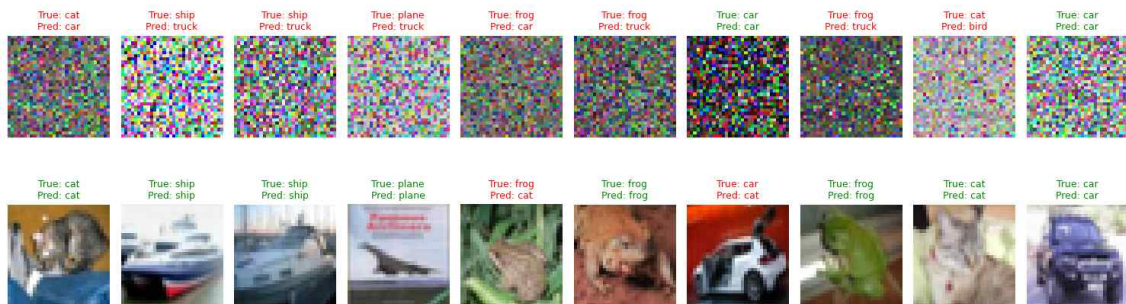
[Original] 테스트 데이터셋 Loss: 1.0780, 정확도: 71.71%

[Randomized] 테스트 데이터셋 Loss: 7.1651, 정확도: 12.62%

원본이미지를 학습한 CNN은 정확도가 FC에 비해 비약적으로 향상되었고, shuffled 이미지를 학습한 FC보다도 성능이 향상되었음.

(53.68% < 71.71%, original FC < original CNN)

Predictions by Model Trained on Original Data



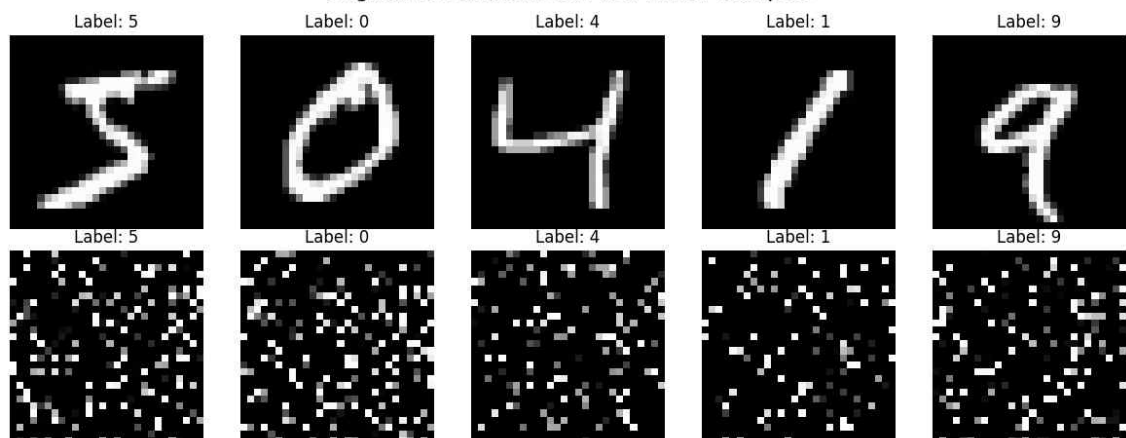
순위를 나열하면 다음과 같음

shuffled CNN << shuffled FC < original FC <<< original CNN

45.31% << 52.71% < 53.68% <<< 71.71%

cifar10에 대한 모델의 정확도가 너무 낮았기에 fc로만으로도 좋은 성능을 낼 수 있는 MNIST데이터셋에 대해서도 학습을 진행하였음.

Original vs Randomized Pixel MNIST Samples



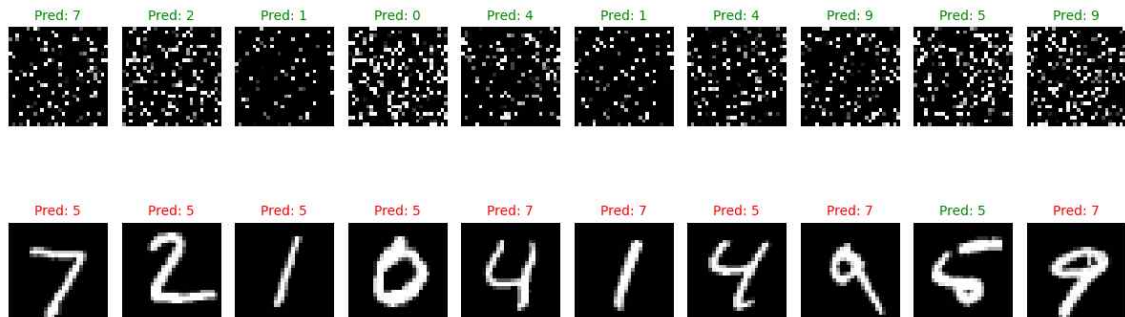
위 이미지는 랜덤하게 셔플된 mnist 데이터셋 이미지 위 그림이 원본. 아래 그림은 shuffled.



## 랜덤 셔플 FC

[Original] 테스트 데이터셋 Loss: 17.3555, 정확도: 6.89%

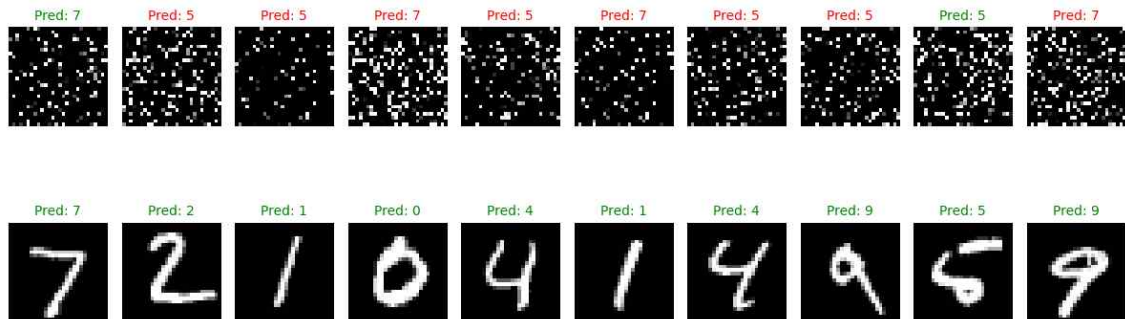
[Randomized] 테스트 데이터셋 Loss: 0.0695, 정확도: 97.77%



## 원본 이미지 FC

[Original] 테스트 데이터셋 Loss: 0.0674, 정확도: 97.94%

[Randomized] 테스트 데이터셋 Loss: 16.4049, 정확도: 10.26%



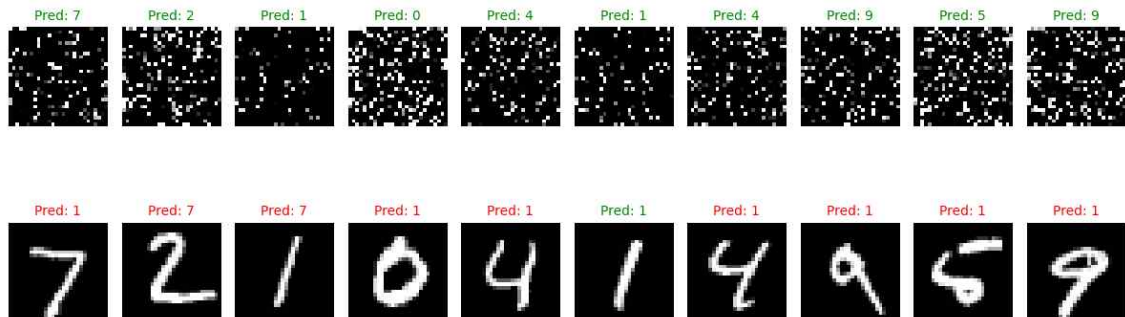
cifar10과 동일하게 두 모델의 정확도 차이가 거의 없음을 확인할 수 있음.

(여기서도 근소하게나마 원본이미지 FC 모델이 성능이 좋았음, 97.77% < 97.94%)

## 랜덤 셔플 CNN

[Original] 테스트 데이터셋 Loss: 8.8015, 정확도: 9.08%

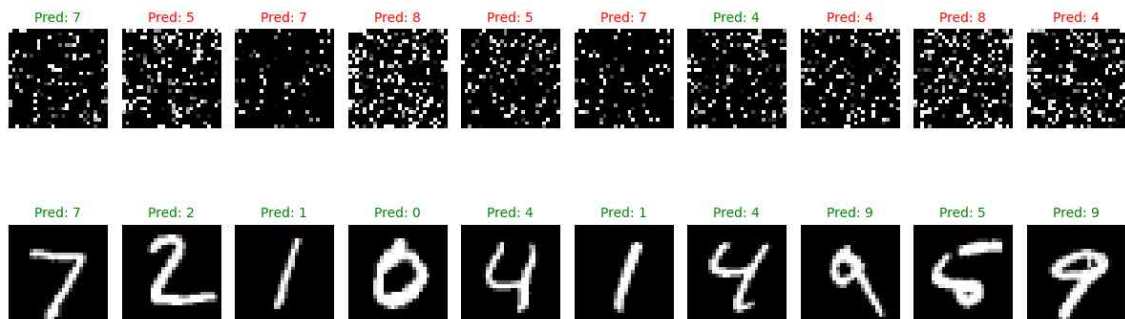
[Randomized] 테스트 데이터셋 Loss: 0.0997, 정확도: 97.58%



## 일본 이미지 CNN

[Original] 테스트 데이터셋 Loss: 0.0259, 정확도: 99.17%

[Randomized] 테스트 데이터셋 Loss: 4.7591, 정확도: 15.65%



cifar10과 동일한 결과를 보임.

shuffled CNN < shuffled FC < original FC << original CNN

97.58% < 97.77% < 97.94% << 99.17%

## 결론:

shuffled image를 학습하던, original image를 학습하던 FC로만 이루어진 Network는 성능의 큰 차이가 없다. 인간이 물체를 인식하는 방식과 굉장히 큰 차이를 지닌다.

shuffled image를 학습한 CNN의 경우 성능이 단순 FC로만 이루어진 Network와 비교하여도 성능이 떨어짐을 확인할 수 있었다. original image를 학습한 CNN의 경우 성능이 비약적으로 향상됨을 확인할 수 있었다.

하지만 shuffled image에 대해서는 CNN이 학습을 못 할 것이라는 예상과 다르게 어느정도 학습됨을 확인할 수 있었다. 이는 CNN으로 이루어진 Network도 인간과 같은 방식으로 이미지를 구분하는 것이 아님을 몸소 느낄 수 있었다. 하지만 FC와 다르게 CNN이 보다 인간의 관점에서 이미지를 분류하는 결과와 비슷함을 확인할 수 있다.