캡스톤 디자인
‘딥페이크 탐지’

#8. Adversarial training

김지수, 김민지, 민지민

지난 캡스톤 회의 내용 -블랙박스 공격 조사-

denoise, sharpening, gaussian noise, salt and pepper, jpeg 으로 만들어진 데이터셋에 대한 강도별 성능 측정

=> 이번주 캡스톤 내용

성능이 잘 나오지 않은 sharpening, gaussian noise, salt and pepper 를 이용해서 adversarial training 진행

데이콘 1등 코드 오류난 부분 코랩으로 실행

→ 또 다른 import 오류로 실행 불가

Adversarial training

```
model = xception(num_out_classes=2, dropout=0.5)
# model = models.xception()
print("=> creating model '{}'".format('xception'))
# model = model.cuda(args.gpu)

# fn = 'deepfake_c0_xception.pkl'
# assert os.path.isfile(fn), 'wrong path'

# model.load_state_dict(torch.load(fn))
best_model = torch.load("/content/drive/My Drive/Anti-forensics/deepfake_c0_xception_tuned.pth.tar")
model.load_state_dict(best_model['state_dict'])
print("=> model weight best_model is loaded")

model.to(device)

criterion = nn.CrossEntropyLoss().cuda()
optimizer = optim.Adam(model.parameters(), lr=0.001, betas=(0.9, 0.999), eps=1e-08)
scheduler = lr_scheduler.StepLR(optimizer, step_size=5, gamma=0.5)
```

저번주에 학습한 모델 가져와
adversarial training 진행

	salt_pepper (strong)	**salt_pepper (medium)**	**salt_pepper (weak)**	**sharpening (strong)**	**sharpening (medium)**	**sharpening (weak)**
gaussian noise strong	loss:5.4151 **accuracy:0.246**	loss: 6.9820 **accuracy: 0.245**	loss:4.5988 **accuracy: 0.263**	loss: accuracy:	loss: accuracy:	loss: accuracy:
gaussian noise med	loss: 8.1304 **accuracy:0.291**	loss: 10.0488 **accuracy: 0.291**	loss : 6.2867 **accuracy : 0.297**	loss: 0.0779 **accuracy: 0.968**	loss: 0.2596 **accuracy: 0.901**	loss: 0.5444 **accuracy: 0.809**
gaussian noise weak	loss: 0.0615 **accuracy: 0.988**	loss:1.2993 **accuracy:0.542**	loss: 3.0358 **accuracy: 0.450**	loss: 0.0358 **accuracy: 0.991**	loss: 0.1095 **accuracy: 0.963**	loss: 0.4228 **accuracy: 0.881**

	sharpening (weak)	**sharpening (medium)**	**sharpening (strong)**	**gaussian noise (weak)**	**gaussian noise (medium)**	**gaussian noise (strong)**
salt & pepper noise (weak)	loss - 8.2891 **acc - 0.255**	loss - 9.0158 **acc - 0.251**	loss -10.5005 **acc - 0.250**	loss - 11.8990 **acc - 0.250**	loss - 12.4867 **acc - 0.250**	loss - 2.7501 **acc - 0.442**
salt & pepper noise (medium)	loss - 0.2123 **acc - 0.912**	loss - 0.1200 **acc - 0.949**	loss - 0.1163 **acc - 0.950**	loss - 0.2564 **acc - 0.890**	loss - 0.1162 **acc - 0.950**	loss - 0.1163 **acc - 0.950**
salt & pepper noise (strong)	loss - 0.0899 **acc - 0.964**	loss - 0.0289 **acc - 0.989**	loss - 0.0280 **acc - 0.989**	loss - 5.1495 **acc - 0.312**	loss - 1.5534 **acc - 0.380**	loss - 0.0276 **acc - 0.989**

	salt & pepper (weak)	salt & pepper (medium)	salt & pepper (strong)	gaussian noise (weak)	gaussian noise (medium)	gaussian noise (strong)
sharpening (weak)	acc - 0.005	acc - 0.017	acc - 0.135	acc - 0.677	acc - 0.936	acc - 1.0
sharpening (medium)	acc - 0	acc - 0	acc - 0.010	acc - 0.011	acc - 0.057	acc - 1.0
sharpening (strong)	acc - 0.069	acc - 0	acc - 0	acc - 0.506	acc - 0.625	acc - 1.0