

영상, 소리데이터를 이용한 자동차 분류기 개발

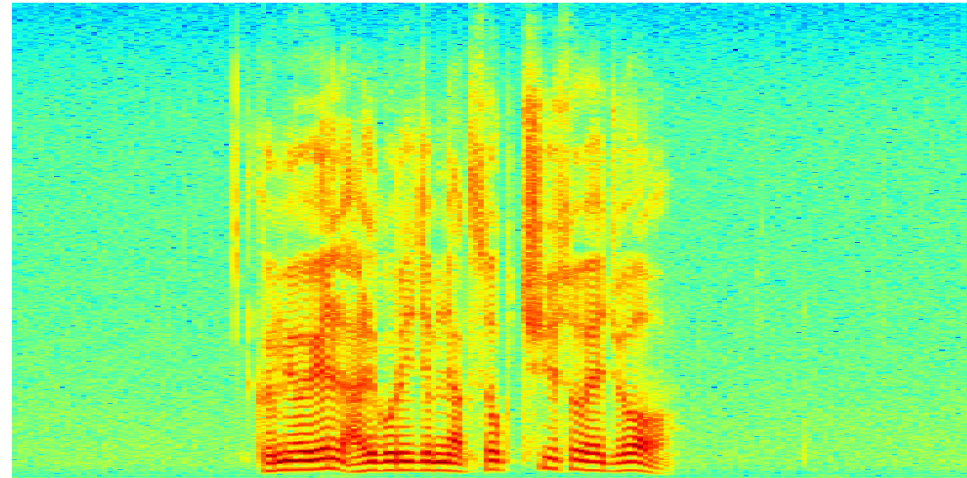
Sogang University
Dept. of Computer Science and Engineering

음성언어연구실

박호성

1. 연구 목표

- 멀티미디어 내의 음성과 정지영상을 input으로 하여 자동차의 종류를 인식하는 프로그램 개발 (5개 classes)
- 소리의 경우 spectrogram으로 변환하여 영상 데이터로 전환한 뒤 사용함



2. 연구 방법론

■ CNN을 활용한 모델 학습

▶ Convolutional Neural Network 사용

- Dataset의 양이 충분하기 때문에 사용 가능 (약 200만 장, class 별 40만)
 - Sound data와 picture data를 조합하여 pair 구성



▶ Spatial Pyramid Pooling Network(SPP)

▶ Inception Network(GoogLeNet)

▶ 현재 SPP와 Inception의 아이디어를 조합하여 사용

- 데이터의 feature가 상대적으로 명확하지 않다고 판단하여 완전한 GoogLeNet으로 구성하지 않음
- 결과: 99.5% accuracy
 - Other class를 추가하면 accuracy 감소할 것으로 보임

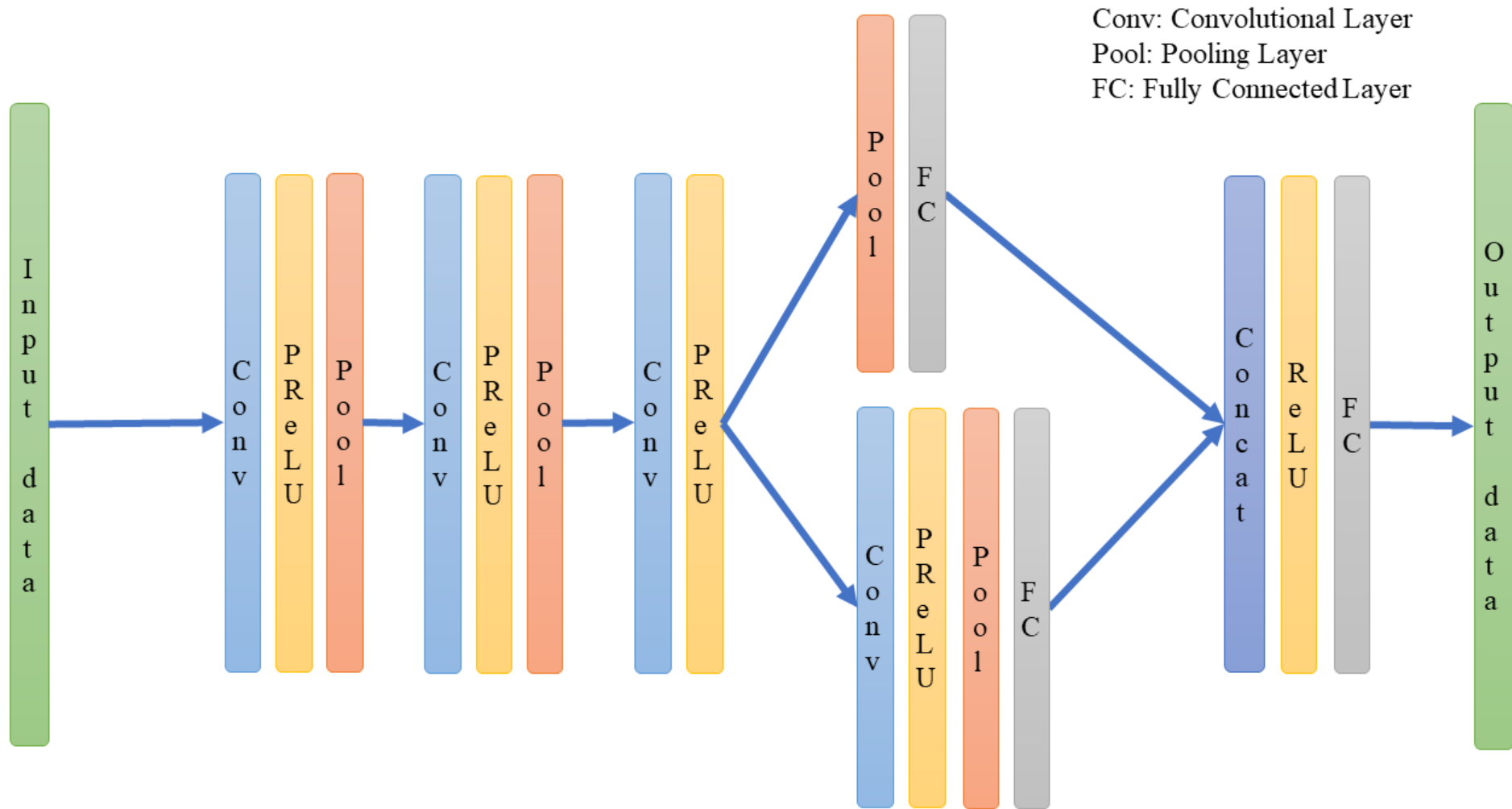
2. 연구 방법론

■ Client

- ▶ AlexNet과 유사한 구조를 가지는 conv layer를 구성하고
- ▶ Inception Network를 참고하여 분기점을 나누어 구성
- ▶ 분기점의 output 을 concat하여 DNN의 input구성

*Inception과 SPP에서 아이디어를 얻었으나, VDCNN을 구성하지 않은 것은 개체 간 feature가 명확하게 드러나지 않기 때문

2. 연구 방법론



2. 연구 방법론

■ Client

- ▶ openCV를 활용한 Windows용 application 개발
- ▶ Caffe module을 활용하여 실시간 classification
- ▶ 학습된 모델을 이용하여 application에 적용

