

영상, 소리데이터를 이용한 차종 분류기 개발

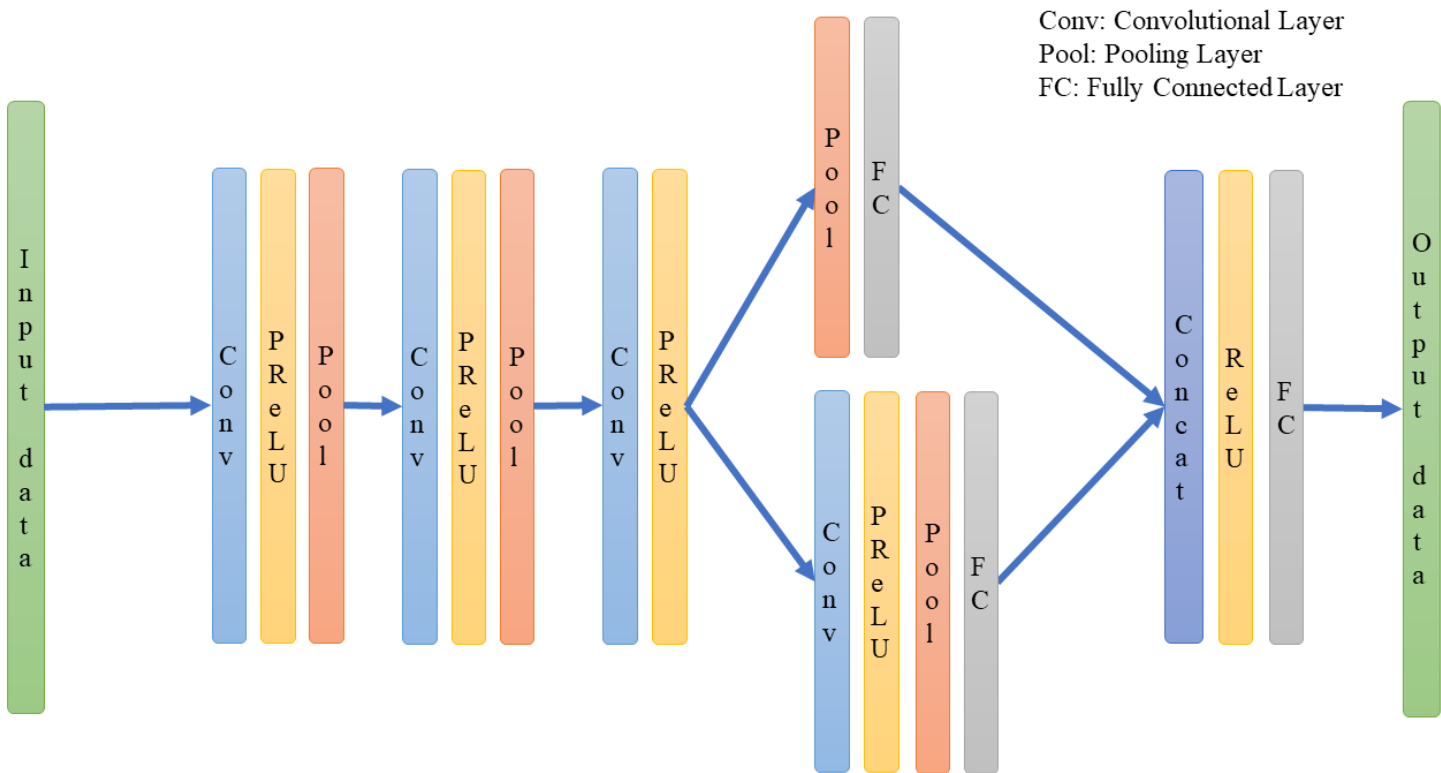
Sogang University
Dept. of Computer Science and Engineering

음성언어연구실

박호성

1. 프로젝트 성과

- 멀티미디어 내의 음성과 정지영상을 input으로 하여 자동차의 종류를 인식하는 프로그램 개발
 - ▶ Truck, Van, Sedan, Bus 및 ETC class (dummy class)
 - ▶ CNN 구조를 Caffe를 사용하여 구현함



1. 프로젝트 성과

■ 프로그램 Input

- ▶ 차량의 측면 사진
- ▶ 차량 소리의 spectrogram



■ 프로그램 output

- ▶ 차량 class의 N-best 결과

1. 프로젝트 성과

■ Dummy Class 구현

- ▶ 영상과 소리를 혼재하여 넣었을 때를 대비한 dummy class 구현
 - Ex> truck 사진과 sedan 소리
- ▶ 일반적인 배경이나 object 사진 등을 삽입
- ▶ 문제점: etc class로 weight값이 집중됨
 - 각 클래스별 적용을 위해 1:1:1:1:4로 데이터를 구성
 - False positive 발생
 - 대부분 Dummy class로 이동
 - 1:1:1:1:1로 구성
 - False negative 발생
 - Dummy로 가야 할 데이터가 가지 않음

1. 프로젝트 성과

- ▶ 데이터 비율에 따라 class 별 score가 집중되기 때문에, 적정 데이터 비율을 설정하던가, 다른 학습방식을 적용해야 함
- **Fine tuning 실시**
 - ▶ 1:1:1:1:4 형태를 유지
 - ▶ 4개 class에 대하여 DNN학습을 실시한 뒤, 5개 class에 대하여 learning rate를 1/100으로 낮추어 finetuning 실시
 - ▶ ETC class로 집중 되는 것을 방지

1. 프로젝트 성과

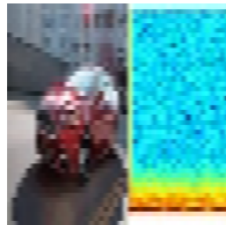
- 프로그램 output
 - ▶ 차량 class의 N-best 결과
 - ▶ Webpage demo 구현

Multimodal Vehicle Classifier

CNN을 활용한 자동차 차종 분류기입니다.

Classification

[Click for a Quick Example](#)



Maximally specific

truck

etc

van

sedan

bus

CNN took 6.488 seconds.

upload vehicle images:

선택된 파일 없음

upload spectrogram images:

선택된 파일 없음

© Sogang University



3. 결론

■ 결론

- ▶ Multi-modal을 통한 객체 분류 성능 향상
 - 영상과의 pair를 통한 학습 데이터량 증대 가능
 - Ex> bus의 sound training data는 30개에 불과하지만, image와 pair 구성함으로써 이미지 1300여개와 조합하여 1300*30개의 가지수로 학습 데이터를 구성 가능
- ▶ Multi-modal을 통한 객체 분류 성능 향상
 - 기존 논문 대비 성능 향상을 보임

	HOG+SVM[2]	CNN
Image	80.39%	93.98%
Image+audio	88.24%	99.5%

3. 결론

■ 결론

- ▶ Very Deep 하지 않은 network에서도 Inception network 기반의 다양한 크기의 필터를 활용한 특징 추출 가능
 - 기존 inception network는 10여개의 convolution layer로 이루어져 있음

REFERENCE

- ▶ [1]Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015, June). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9).
- ▶ [2]Wang, T. & Zhu, Z. (2012, January). Real time vehicle detection and reconstruction for improving classification. In IEEE Computer Society's Workshop on Applications of Computer Vision (pp. 9-11).
- ▶ [3]He, K., Zhang, X., Ren, S., & Sun, J. (2014, September). Spatial pyramid pooling in deep convolutional networks for visual recognition. In Proceedings of the European Conference on Computer Vision (pp. 346-361).
- ▶ [4]Simonyan, K. & Zisserman, A. (2014, December). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.