PORTFOLIO

GISU CHUNG
정지수

<contents>

1. 학부 연구 이력

(1). 'Sift-Nonsift classification of image patches using CNN' 주제로 연구

한국방송미디어 공학회 공모(포스터 발표) 제 2 저자, 장려상 수상

Abstract: 본 연구는 주어진 영상 패치가 SIFT 특성(즉 SIFT 특징점)을 포함하는지 아니면 SIFT 특성을 나타내지 않은지(즉, NonSIFT)를 구분하기 위해 CNN 구조를 활용한다. 우선 CNN 을 훈련시키기 위해 기존의 SIFT 알고리즘으로 SIFT 특징점을 검출하고 이 점들을 중심으로 SIFT 영상 패치를 생성하며, 반대로 SIFT 특징점을 포함하지 않는 영상 패치를 같은 수만큼 생성하여 데이터 베이스를 구축한다. 이를 바탕으로 3 개의 컨벌루션 레이어를 갖는 CNN 구조를 훈련시켰다. 훈련된 CNN 에 대해 오픈소스 라이브러리 VL_FEAT 의 SIFT 알고리즘을 수행한 결과를 그라운드 트루스 (ground truth)로 정하고 제안된 방법의 결과와 비교한 결과 90.8%의 분류 정확도를 얻었으며 CNN 을 이용한 SIFT 특징점 검출의 가능성을 확인하였다.

연구내용: SIFT(Scale-Invariant Feature Transform)는 영상의 특징점을 추출하는 전통적인 영상처리 기법으로, 카메라 캘리브레이션, 객체 추출, 영상 스티칭 등에 활용된다. 이러한 특징점 추출 방법을 CNN 이 학습 가능한지에 대한 의문을 해결하고자 MS COCO dataset 으로부터 각 이미지별로 SIFT 특징점을 추출, descriptor 크기의 영상패치를 추출했다. SIFT 성분이 없는 순수한 Non-SIFT 패치 추출을 위해 모든 SIFT 특징점이 잡힌 부분에 descriptor 크기의 가중치를 더해 saliency map 을 제작했다. 이를 통해 SIFT 주변 영역으로부터 완전히 분리된 영역에서 순수한 NonSIFT 영역을 확보했다. 확보된 데이터 세트로부터 얕은 CNN 을 훈련시켜 inference time 을 줄이면서 해당패치가 SIFT 특징을 가졌는지, 가지지 않았는지에 대해 90.8%의 성능으로 분류하는 모델을 개발했다.

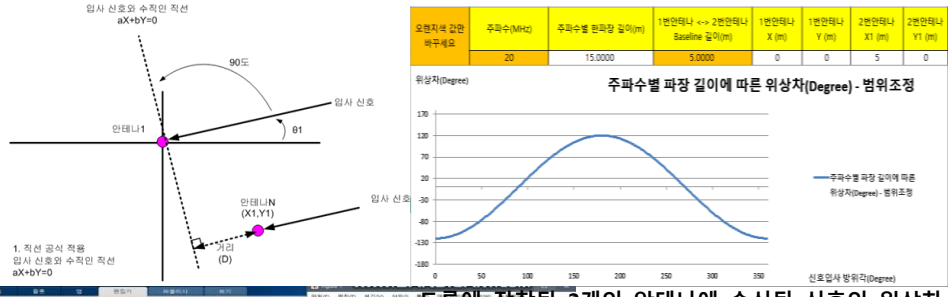
〈그림 1〉 SIFT saliency map (왼쪽), 원영상 (오른쪽)

〈그림 2〉 32x32 SIFT 및 NonSIFT 영상 패치

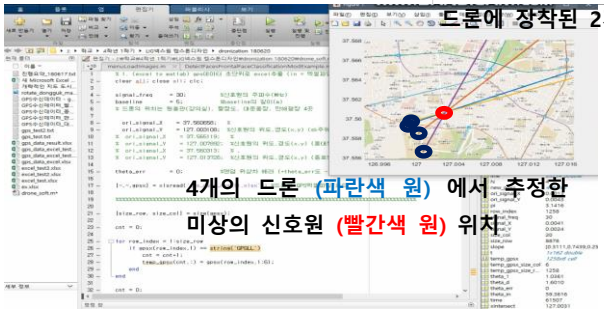
(2). LIG 넥스원 전자전연구소 공모 '드론을 이용한 미상의 신호원 탐지' 프로젝트

캡스톤 디자인 기업사회 연계 프로젝트, 우수상 수상

2 개 이상의 수신 안테나에 대해 주파수별 파장에 따른 위상차를 계산하고, 이를 통해 역산된 입사방위각으로부터 2 개의 드론을 가정해 신호원 까지의 line 을 그려 교점 위치를 확정하였다. ground truth 인 수신된 GPS 좌표로부터 수신된 위치와의 비교를 통해 정확도를 측정, 98%의 성능을 보였다. GPS 좌표는 드론에 탑재한 GPS 수신기로부터 얻은 신호를 무선 시리얼 컨버터를 통해 전달받도록 하드웨어를 구성했다.



드론에 장착된 2개의 안테나에 수신된 신호의 위상차



4개의 드론 (파란색 원) 에서 추정한 미상의 신호원 (빨간색 원) 위치

2. 석사과정 연구 이력(1)

(1). 'Filter pruning by image channel reduction in pre-trained convolutional neural networks' 연구

SCI 급 저널 MTAP(Multimedia Tools and applications) 1 저자 논문게재 및 특허등록

Abstract: There are domain-specific image classification problems such as facial emotion and house number classifications, where the color information in the images may not be crucial for recognition. This motivates us to convert RGB images to gray-scale ones with a single Y channel to be fed into the pre-trained convolutional neural networks (CNN). Now, since the existing CNN models are pre-trained by three-channel color images, one can expect that some trained filters are more sensitive to colors than brightness. Therefore, adopting the single channel gray-scale images as inputs, we can prune out some of the convolutional filters in the first layer of the pre-trained CNN. This first-layer pruning greatly facilitates the filter compression of the subsequent convolutional layers. Now, the pre-trained CNN with the compressed filters is fine-tuned with the single-channel images for a domain-specific dataset. Experimental results on the facial emotion and Street View House Numbers (SVHN) datasets show that we can achieve a significant compression of the pre-trained CNN filters by the proposed method. For example, compared with the fine-tuned VGG-16 model by color images, we can save 10.538 GFLOPs computations, while keeping the classification accuracy around 84% for the facial emotion RAF-DB dataset.

연구내용: Deep learning 을 적용할 때, 대부분 ImageNet(color images) based 로 훈련된 네트워크가 사용됨에 주목, 사람 얼굴의 감정 인식을 위한 CNN 은 이미지의 칼라성분보다 밝기성분이 더 중요할 것이라고 생각. 이에 따라 우선 감정인식 task 데이터베이스인 RAF-DB 에 대해 각각 칼라와 흑백 이미지로 전이학습한 두 모델의 성능 비교하였고 성능은 1% 정도의 미미한 차이가 있었음. 따라서 감정인식과 같은 task 를 훈련 시 칼라로 사전 훈련된 네트워크에 대해서 내부의 학습된 필터에 칼라 고유의 중복성이 함께 존재할 것이라 가정, taylor-expansion based filter pruning 방법과 sparse structure selection pruning 기법에 대해 1 채널 흑백 영상으로 변환 후 적용했을 때 압축 효율이 증가한다는 것을 정성적/정량적으로 보임.

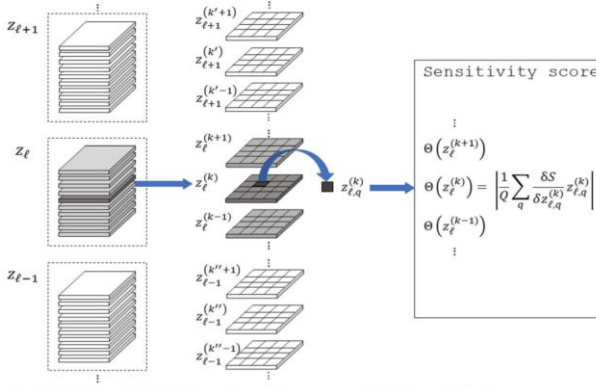


Fig. 2 Calculations of sensitivity scores for each feature map (activation output) in Eq. (4)

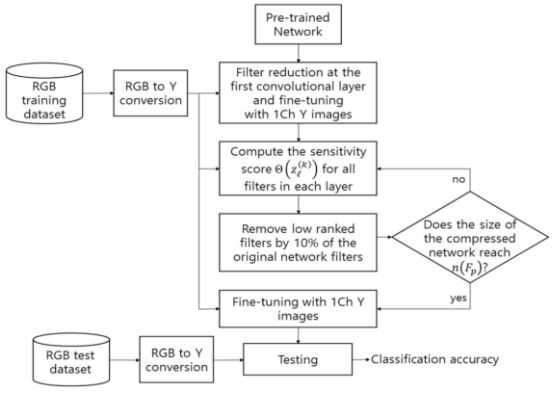


Fig. 3 Block diagram of the all-layer filter compression

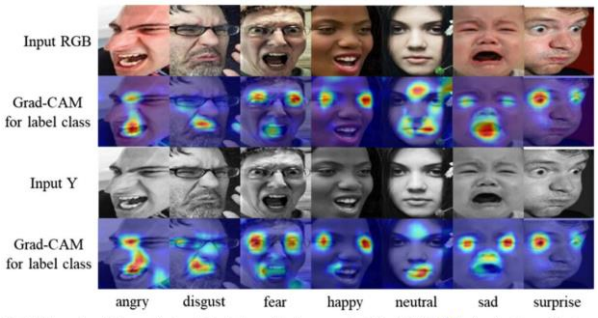


Fig. 1 Examples of the gradient-weighted class activation mapping (Grad-CAM) [21] visualizations.: Top two rows are RGB images and the corresponding Grad-CAM from the RGB-trained VGG-16 [22] and the bottom two rows are Y images and the corresponding Grad-CAM from the Y-trained VGG-16

Table 4 Comparisons of classification accuracies between 3Ch RGB (all-layer compressions without the first-layer pruning) and 1Ch Y (all-layer compressions after the first-layer pruning) models for VGG-16 [22] and RAF-DB [15] datasets

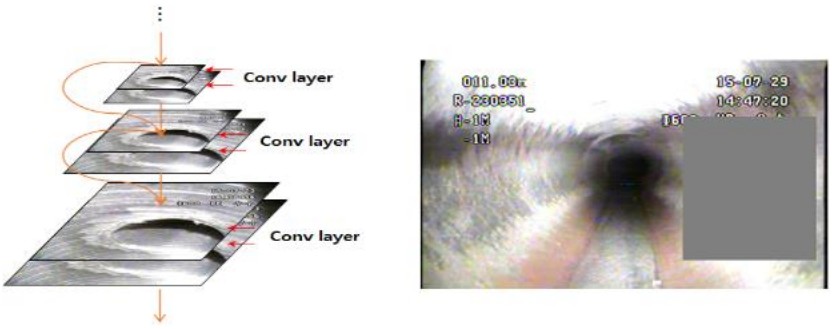
Target percentage of filter pruning	3Ch RGB			1Ch Y		
	Accuracy	Processing time	# FLOPs	Accuracy	Processing time	# FLOPs
40%	84.84%	12.37 ms	6.970G	85.36%	11.81 ms	6.545G
50%	83.83%	10.86 ms	5.301G	85.17%	10.07 ms	5.003G
60%	83.14%	8.48 ms	3.751G	83.51%	7.66 ms	3.462G
70%	81.78%	6.43 ms	2.448G	82.37%	5.91 ms	2.058G
80%	77.51%	5.23 ms	1.571G	79.66%	4.93 ms	1.297G
90%	54.95%	4.08 ms	415.274 M	56.12%	3.98 ms	373.043 M

(2). ‘Deep learning based model for detecting sewer pipe defects’ 연구

제 1 발명자 특허등록, 대한토목공학회 공모 1 저자 논문게재(구두발표)

Abstract: 하수도관 내부 결함은 지반침하의 잠재원인으로서 정밀히 조사하여 그 정도에 따라 하수관을 교체하거나 부분적 보수를 통해 심각한 사고를 사전에 예방해야 한다. 본 연구에서는 최근 다양한 분야에서 활용되고 있는 인공지능을 기반으로 한 하수도관 결함 탐지 모델을 개발한다. 또한 분류 성능 향상을 위해 피라미드형태의 구조를 사용하여 하수도관 이미지를 여러 스케일로 분할하여 특징을 추출하도록 하고, 컷아웃 기법을 적용해 네트워크가 일부 차별성 있는 특징만을 학습하는 것을 방지한다. 실험결과의 비교를 통해 제안한 방법의 타당성을 보이고, 정상 라벨을 포함한 하수도관 내에 발생할 수 있는 모든 결함 항목에 대해 89.61% 의 정확성으로 실제 하수도관 정밀검사시 유용하게 사용될 수 있음을 보인다.

(a) 피라미드 네트워크 구조 (b) Cutout augmentation



연구내용: 건설 토목회사 ㈜해문개발에서 인턴 기간동안 수행, 회사 내 사업 부문 중 하나인 이상 하수도관 결함 탐지에 대해 선행연구 리서치 후 아이디어 제안함. 당사에서 보유한 하수도관 탐지 비디오에 대해 라벨링 작업 후 항목별 이미지 추출, 네트워크 훈련을 통해 80% 이상의 성능으로 AI 로 탐지를 대체할 가능성을 열고 특허를 등록함. 또한 Pyramid network 의 도입과 Cutout augmentation을 통해 성능을 약 9% 증가시켜 관련한 내용을 논문으로 작성해 대한토목공학회에 발표함. 이후 모델에 pruning 과 quantization 을 적용해 학습된 모델의 경량화를 진행하였고, NVIDIA Jetson nano 키트에 모델을 탑재하여 CCTV 로봇에 부착, 실제 운용할 때 모델을 real-time inference 하여 main server 에 전달, 결함 이상 항목의 가능성이 높은 곳을 정밀히 촬영할 수 있게 기술을 향상시킴.

3. 석사과정 연구 이력(2)

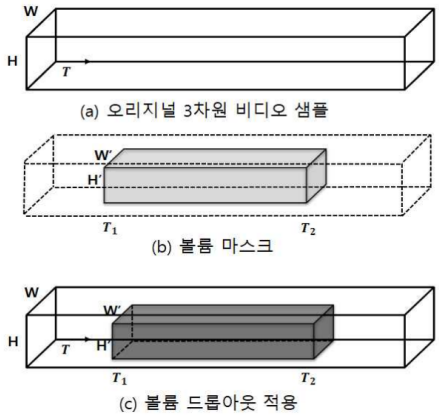
(1). 'Volume dropout on 3D convolutional neural network for video action recognition'

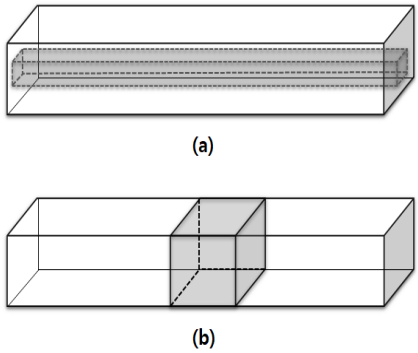
석사 학위논문 관련 연구

Abstract: The number of labeled videos is often insufficient to train lots of parameters associated with deep 3D CNN. Therefore, to diversify the given labeled videos and to alleviate the network overfitting, video augmentation is necessary in training 3D CNN. Meanwhile, when using 2D CNN, which is trained by 2D images has proved that the image augmentation using regional dropout makes the network to focus on less discriminative areas of the image and to alleviate the overfitting problem. This motivates us to extend the spatial dropout in 2D images into the 3D volume dropout for videos. In this paper, we propose a novel video augmentation technique for video action recognition with using 3D CNN.

연구내용: 기존의 2D deep learning 분야는 부족한 데이터로 인한 낮은 성능을 개선시키기 위한 다양한 데이터 증강방법이 제시되었으나, 3D 비디오에 대해선 연구가 이루어지지 않음. 따라서 비디오에 적용할 수 있는 증강 방법에 대한 연구를 진행, 기존 이미지에 적용시 뛰어난 성능을 보인 Cutout 증강 방법을 발전시켜 비디오에 적용할 수 있는 볼륨 드롭아웃 기법을 제안함. 아래 그림과 같이 볼륨 마스크를 샘플링하고, 원본 샘플과의 element-wise 곱을 통해 데이터를 다양화할 수 있음. 비디오는 시간축이 존재하므로 공간적 드롭아웃(spatial dropout)과, 시간적 드롭아웃 방법(temporal dropout) 두 가지 방법을 제안하였음. 각 방법을 통해 two-stream 3D CNN에 대해 HMDB-51 데이터베이스로 훈련했을 때, 최대 1.75%의 성능 향상을 보임.

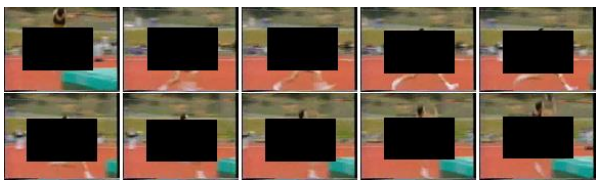
[그림 3-1] 비디오 데이터 볼륨 드롭아웃



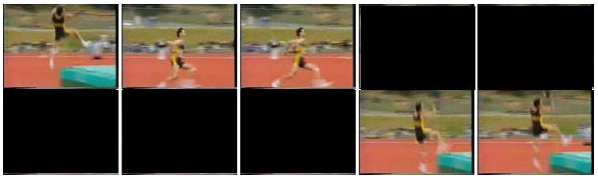


Volume dropout
(a) spatial dropout 적용
(b) temporal dropout 적용
흑백 영역이 dropout 영역

(a) original sample



(b) After spatial-wise volume dropout



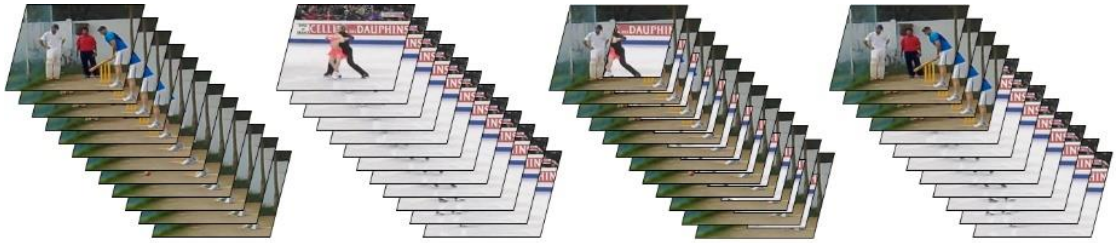
(a) After temporal-wise volumedropout

(2). 'StackMix and TubeMix : Video augmentation strategy for action recognition' 연구

Abstract: The number of labeled videos is often insufficient to train lots of parameters associated with deep 3D CNN (Convolutional Neural Network). Therefore, to diversify the given labeled videos and to alleviate the network overfitting, a video augmentation is necessary in training 3D CNNs. Meanwhile, deep CNNs trained by 2D images have already proven that the image augmentation using regional dropout makes the network to focus on less discriminative areas of the image and to alleviate the overfitting problem. That is, many studies have shown that the augmentation method of mixing two images and their ground truth enhances the performance of 2D CNNs. This motivates us to extend the spatial dropout in 2D images into the volumetic dropout for videos. Since the video data have two domains of space and time, we can extend the dropout into the two domains, both spatially and temporally. Specifically, we propose novel augmentation strategies, Tubemix and Stackmix. The Tubemix is to mix two videos samples in a video-in-video fashion to enjoy the spatial dropout effect. Similarly, the Stackmix is to combine two videos in a video-to-video fashion to have a temporal dropout effect. Experimental results on the UCF-101 and HMDB-51 datasets show the accuracy improvements around 1.45% and 1.24%, respectively, with the proposed video augmentations.

연구내용: 학위논문 연구를 기반으로 두개 이상의 복수 샘플 데이터를 섞는 기존의 2D 증강방법의 트렌드를 따라 비디오에 새롭게 적용시킴. Volume dropout 기법에선 단순히 dropout 시킨 영역을 다른 비디오로 대체함. 즉, 비디오 내에 또다른 비디오가 삽입된 Tubemix 와, 비디오에 또다른 비디오가 이어붙어진 Stackmix 기법을 제안함.

기존의 image augmentation 기법으로 큰 성능 향상을 보인 CutMix 방법.



제안한 방법 적용 결과