

빅데이터로부터 추출된 주변 환경 컨텍스트를 반영한 딥러닝 기반 거리 안전도 점수 예측 모델

이기인[†], 강행봉^{**}

A Deep Learning-based Streetscapes Safety Score Prediction Model using Environmental Context from Big Data

Gi-In Lee[†], Hang-Bong Kang^{**}

ABSTRACT

Since the mitigation of fear of crime significantly enhances the consumptions in a city, studies focusing on urban safety analysis have received much attention as means of revitalizing the local economy. In addition, with the development of computer vision and machine learning technologies, efficient and automated analysis methods have been developed. Previous studies have used global features to predict the safety of cities, yet this method has limited ability in accurately predicting abstract information such as safety assessments. Therefore we used a Convolutional Context Neural Network (CCNN) that considered "context" as a decision criterion to accurately predict safety of cities. CCNN model is constructed by combining a stacked auto encoder with a fully connected network to find the context and use it in the CNN model to predict the score. We analyzed the RMSE and correlation of SVR, Alexnet, and Sharing models to compare with the performance of CCNN model. Our results indicate that our model has much better RMSE and Pearson/Spearman correlation coefficient.

Key words: Urban Safety, Crime Prediction, Stacked Autoencoder, Context

1. 서 론

지난 수십 년간 도시가 속한 지역경제를 활성화시키기 위해 다양한 연구들이 이루어지고 있다. 이러한 연구들은 야간 경제(NTE: Night-Time Economy) 개념의 창조를 통해 도시를 즐거움과 소비의 공간으로 만드는 시도[1, 2], 지역축제를 개최함으로써 경제적 효과를 얻으려는 시도[3] 등 관광객을 끌어들일 수 있는 매력적인 도시를 만드는데 중점을 두고 있

다. 이러한 가운데 안전한 공간을 창출하면 방문객수와 소비자 지출에 긍정적인 영향을 주는 것으로 알려져 범죄에 대한 공포감의 완화가 지역경제 활성화에 중요한 역할을 맡게 되었다[4]. 따라서, 범죄에 대한 공포감을 감소시키기 위해서 범죄 발생 혹은 안전도 예측에 대한 연구가 진행되고 있다. 과거에는 이러한 분석에 수많은 정보와 비용이 필요했지만, 최근 컴퓨터 비전과 빅데이터가 발전함에 따라 영상으로부터 도시 환경을 예측하는 것이 가능해졌고, 특히

※ Corresponding Author : Hang-Bong Kang, Address: (14662) 43, Jibong-ro, Bucheon-si, Gyeonggi-do, Republic of Korea, TEL : +82-2-2164-4598, FAX : +82-2-2164-4945, E-mail : hbkang@catholic.ac.kr

Receipt date : July 19, 2017, Approval date : July 31, 2017

[†] Dept. of Media Technology Contents., The Catholic University of Korea

(E-mail : marko0332@catholic.ac.kr)

^{**} Dept. of Media Technology Contents., The Catholic University of Korea

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (No. 2015R1A2A1A10056304).

범죄 발생 예측에 대한 Bachner 등[5]의 연구와 Thompson 등[6]의 연구는 사회적으로 큰 관심을 받고 있다.

도시의 안전도를 예측하는데 있어 Wilson과 Kelling의 '깨진 창문 이론[7]'을 바탕으로 하는 방법이 있다. 깨진 창문 이론이란 거리에 깨진 유리창 하나를 방치해 두면 그 지점을 중심으로 범죄가 확산되기 시작한다는 이론으로, 사소한 무질서를 방치했을 경우 도시 전체의 혼란으로 퍼지게 된다는 것이다. 다시 말해 도시에서 시각적으로 무질서하게 느껴지는 장소는 실제로도 낮은 안전도를 가질 가능성이 높다고 판단할 수 있다. 이 이론을 토대로 시각적 인지(Visual Perception)를 이용해서 도시 거리 영상에 대한 안전도를 평가하고, 그 값들을 기계학습 알고리즘을 통해 학습하고 예측하는 연구들이 이루어졌다[8-14].

대부분의 기존 연구들은 도시 영상에서 전역 특징 추출 방법을 통해 색상, 그레디언트, 텍스처 등의 저수준 특징(Low Level Feature)을 얻었다. 때문에 도시 환경의 안전도 평가와 같은 추상적인 정보를 예측하는데 있어 한계가 존재한다. 이를 해결하기 위해서 본 논문에서는 영상으로부터 추상적인 고수준 특징(High Level Feature)을 추출하는데 뛰어난 합성곱(Convolution)과, 비감독 학습(Unsupervised Learning)으로 많이 쓰이는 자기부호화기(Autoencoder)를 활용해서 Saliency map을 얻고[15,16], 시각적 인지를 통해 도시 영상의 안전도를 판단하는데 있어 핵심 역할을 하는 환경적 컨텍스트(Environmental Context)를 찾는다[17]. 데이터 수가 적으면 기존의 지도학습(Supervised Learning) 방법으로는 쉽게 학습 데이터에 특화(Overfitting)될 수 있기 때문에 우리는 누적 합성곱 자기부호화기(SCA: Stacked Convolutional Autoencoder)를 통해 안전/위험에 관련된 컨텍스트를 각각 학습하고, 이 둘을 결합(Joint)하여 도시의 안전도를 예측하는 새로운 모델인 CCNN(Convolutional Context Neural Network)을 제시한다.

본 논문의 구성은 다음과 같다. 2절에서는 본 논문의 관련 연구에 대해서 설명하고, 3절에서는 제안한 CCNN 모델의 전체적인 개요와 도시 영상에 대해 순위를 매기는데 사용한 Trueskill 알고리즘[18]에 대해서 설명한다. 4절에서는 본 논문에서 제안한 모델에 사용하는 Saliency Estimation 알고리즘[15]과

Convolutional Context Neural Network(CCNN)의 구조 및 학습 방법에 대해서 설명한다. 5절에서는 실험 방법과 평균 자승 오차와 상관관계 분석을 이용하여 다른 모델들과의 성능 비교 결과를 보여준다.

2. 관련 연구

범죄 예측 및 도시의 안전도 예측에 대한 다양한 연구들이 이루어지고 있다. Bachner 등[5]은 Netflix와 WalMart와 같은 소매 업체가 개발한 비즈니스 기술을 사용하여 소비자 행동을 예측하고, 이를 토대로 범죄를 예측한다. Thompson 등[6]은 캘리포니아의 산타 크루즈에 적용되어 적은 인원으로도 효과적이고 효율적인 범죄대응을 할 수 있게 도와준 연구로, 빅데이터를 토대로 범죄가 발생할 가능성이 높은 시간대와 장소를 예측하였다. Mohler 등[8]은 지진이 일어난 뒤에 따라오는 여진과 같이 범죄가 최초 발생 지점을 중심으로 확산된다는 것을 발견하였다.

한편, Salesses 등[9]은 시각적 인지를 바탕으로 진행한 쌍대비교실험(Pairwise Comparison)을 통해 도시 환경의 독특함, 부유함과 안전도 점수를 측정하였고, 이를 바탕으로 도시 환경을 분석 및 각각의 범주에 대한 점수 예측 방법을 연구하였다. Ordonez 등[10]은 Salesses 등[9]의 데이터 셋을 확장시켰고, GIST[19], Fisher Vector[20], DeCAF[21]를 사용하여 점수 예측에 필요한 특징을 추출 하였다. 이렇게 얻은 특징들을 통해 서포트 벡터 머신(SVM: Support Vector Machine)과 서포트 벡터 회귀(SVR: Support Vector Regression)을 사용해서 모델을 학습하였다. Khosla 등[11]은 주어진 도시 영상으로부터 특정 목적지를 찾아가는 방법, 목적지까지의 거리 그리고 그 장소의 안전도를 추론하는 연구를 실시하였다. Naik 등[12]은 Salesses 등[9]의 데이터 셋을 이용해서 HoG[22], GIST 등의 특징을 통해 안전도 점수를 예측하는 모델을 제안하였다. Mohler 등[12]은 지진이 일어난 뒤에 따라오는 여진과 같이 범죄가 최초 발생 지점을 중심으로 확산된다는 것을 발견하였다. 이러한 발견을 토대로 과거 범죄 기록을 이용하여 범죄 예측 모델을 제안하였다. Gerber 등[13]과 Chen 등[14]은 범죄 발생과 관련되거나 근접한 지역에서는 SNS에 부정적인 글이 많이 올라온다는 점을 활용하여 범죄 발생을 예측하였다.

기존 연구들은 대부분 세 단계에 걸쳐 진행되는 데, 먼저 학습에 사용할 데이터를 수집한다. 과거 범죄에 대한 기록들과 설문 내용, SNS, 그리고 영상 등을 주로 학습 데이터로 사용하였고, 다음으로 데이터로부터 특징을 추출한다. 범죄 발생 시간과 장소에 대한 특징을 찾거나, 영상을 데이터로 쓸 경우에는 색상, 그래디언트 등의 전역 특징을 추출하였다. 마지막으로 추출한 특징들을 토대로 예측 모델을 생성한다. 하지만 이러한 저수준 특징을 사용해서 안전도 예측과 같은 추상적인 답을 구하는데 한계가 있고, 본 논문에서는 이를 해결하기 위해 판단의 핵심이 되는 컨텍스트를 추출하여 학습에 사용한다.

3. 거리 안전도 측정

3.1 개요

본 논문에서 제안하는 안전도 예측 모델의 흐름은 Fig. 1와 같다. 먼저, 도시 영상에 대해서 쌍대비교실험을 통해 구한 경쟁 결과를 토대로 Trueskill 알고리즘을 사용하여 도시 영상의 순위를 정한다. 그다음 C-S(Center-Surround) 재구성 모델을 통해 Saliency map을 구하고, 구한 Saliency map을 사용하여 컨텍스트를 추출한다. 마지막으로 context map을 신경망에 입력으로 넣어 안전도 점수를 예측한다.

3.2 Trueskill을 이용한 영상 순위

우리는 Place Pulse 1.0 데이터[9]와 우리가 수집한 데이터로 구성된 도시 영상 데이터 셋을 이용하여 모델을 학습하였다. 실험 참가자들에게 구글 스트리트 뷰(Google Streetview)영상을 임의로 보여주고

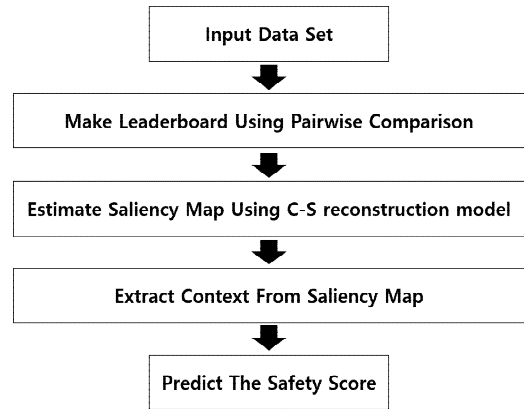


Fig. 1. Flowchart for safety score prediction model algorithm.

‘어느 장소가 더 안전해 보이는가?’라는 질문에 왼쪽, 오른쪽 혹은 동등하다 중 하나를 선택할 수 있게 하였다(Fig. 2-(a)). 2,286장의 도시 영상에 대해 총 32명의 참가자들이 23,361번의 쌍대비교실험을 진행하여 시각적 인지에 따른 안전도를 구하기 위한 데이터를 수집하였다. 실험을 통해 얻은 데이터를 마이크로소프트의 Trueskill 알고리즘[18]을 통해 최종 안전도 점수를 구한 뒤 순위를 매겼고, 각각의 안전도 점수를 0-10의 값으로 정규화 하였다(Fig. 2-(b)).

Trueskill 알고리즘[18]은 온라인 게임에서 플레이어의 순위를 결정하기 위해 베이지안 그래픽 모델을 사용한다. 우리는 주어진 질문에 따라 선택된 영상을 일대일 경쟁에서의 승리자로 간주하였다. 각 영상의 스킬은 $N(\mu, \sigma^2)$ 으로 모델링되며 경쟁이 진행될 때 마다 업데이트 된다. 두 명의 플레이어 x와 y에 대해서 플레이어 x가 y를 이겼을 경우 업데이트는



Fig. 2. (a) To obtain a ranked score using Trueskill, participants were asked to select the images that look more safe. They can select left, right or equal. (b) After pairwise comparisons, the images are ranked on their perceived safety(S_i) between 0 and 10.

Equation (1), (2)를 통해 진행된다.

$$\mu_x^{new} = \mu_x^{old} + \frac{\sigma_x^2}{c} \cdot f\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\epsilon}{c}\right) \quad (1)$$

$$\mu_y^{new} = \mu_y^{old} + \frac{\sigma_y^2}{c} \cdot f\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\epsilon}{c}\right)$$

$$\sigma_x^{new} = \sigma_x^{old} \cdot f\left[1 - \frac{\sigma_x^2}{c^2} \cdot g\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\epsilon}{c}\right)\right] \quad (2)$$

$$\sigma_y^{new} = \sigma_y^{old} \cdot f\left[1 - \frac{\sigma_y^2}{c^2} \cdot g\left(\frac{(\mu_x - \mu_y)}{c}, \frac{\epsilon}{c}\right)\right]$$

$$c^2 = 2\beta^2 + \sigma_x^2 + \sigma_y^2$$

x와 y의 trueskill은 $N(\mu_x, \sigma_x^2), N(\mu_y, \sigma_y^2)$ 이다. β 는 사전 정의된 상수로 게임 별 편차를 의미하며, ϵ 은 경험적으로 추정된 x, y 두 플레이어 가 비길 확률이다. 함수 $f(\theta) = N(\theta) / \Phi(\theta)$ 와 $g(\theta) = f(\theta) \cdot (f(\theta) + \theta)$ 는 정규 확률 밀도 함수 $N(\theta)$ 와 정규 누적 밀도 함수 $\Phi(\theta)$ 를 사용하였다. 우리는 모든 영상에 대해서 $\mu = 25, \sigma = 25/3$ 을 초기값으로 사용하였으며, $\beta = 25/3, \epsilon = 0.1333$ 을 사용하였다[18]. Equation (1)은 플레이어의 평균값을, Equation (2)는 플레이어의 분산 값을 업데이트 한다. Equation (1), (2)를 통해 업데이트된 영상들의 trueskill에 대해서 평균값을 의미하는 μ 를 최종순위를 결정하는데 사용하였다.

4. 컨텍스트 기반 안전도 예측 모델

4.1 컨텍스트 추출 알고리즘

본 논문에서는 Xia 등의 Saliency Estimation 알고리즘 구조[15]를 사용해서 Context Map을 얻는다. Fig. 3-(a)는 Saliency Estimation 알고리즘의 구조를 보여준다. 입력으로 들어온 영상에 대해서 $16 \times 16 \times 3$ 의 크기를 가지는 훈련 영상 8,000개를 임의로 추출하고, 각각의 샘플 영상의 중앙에서 $8 \times 8 \times 3$ 의 크기를 가지는 타겟 영상 8,000개를 추출한다. 이렇게 얻은 훈련 영상과 타겟 영상을 통해 C-S 재구성 모델을 학습시킨다. C-S 재구성 모델은 4개의 인코더 층과 4개의 디코더 층, 그리고 2개의 추론 층으로 이루어져 있다(Fig. 3-(b), (c))

인코더는 컨볼루션과 max-pooling을 두 번 반복하고, 디코더는 max-unpooling과 디컨볼루션을 두 번 반복한다. 추론 층에서는 오토인코더를 통해 나온 영상을 바탕으로 타겟 영상을 추론하는데 완전 연결

층을 사용하며 각각 768개, 192개로 구성된다. 이렇게 구한 뉴런은 출력 크기를 $8 \times 8 \times 3$ 으로 변경하여 타겟 영상과 크기를 일치시킨다. 각 층의 결과는 ReLU와 dropout 기법을 적용하였고, 컨볼루션을 할 때 각각 3×3 과 5×5 크기의 커널을 사용하였다. 학습이 완료된 C-S 재구성 모델을 사용하여 전체 입력 영상에 대해서 예측 영상과 원본 영상의 차이(residual map)를 구한다. 이렇게 구한 Context map을 원본 영상에 곱해서 불필요한 정보는 없애고 판단에 영향을 주는 특징만 남긴다(Fig. 3-(d)).

4.2 CCNN (Convolutional Context Neural Network)

영상을 판단함에 있어 컨텍스트에 따라 결과가 달라질 수 있다는 것[17]을 토대로 Fig. 3-(a)의 모델을 통해 컨텍스트를 추출한다. 우리는 Fig. 2-(a)의 실험을 통해 얻은 결과를 바탕으로 2,286개의 영상을 인지된 안전도에 따라 순위를 매기고, 이 데이터를 토대로 순위 상위 30%와 하위 30%의 영상을 각각 안전한 영상, 위험한 영상으로 판단하였다. 그 후에 안전한 영상과 위험한 영상에서 컨텍스트를 추출하고 이를 각각의 특징 추출 모델을 학습시키는데 사용하였다. (Fig. 4) 특징 추출 층은 2개의 컨볼루션 층과 2개의 max-pooling 층으로 구성하였으며 ReLU, dropout 기법을 사용하였다. 마지막 완전 연결 층에서는 ground truth와 모델이 예상한 score의 차이를 사용하여 손실을 구한다.

학습 모델에 고정된 크기의 영상($210 \times 320 \times 3$)이 입력으로 들어오면 컨텍스트를 추출하고, 추출된 컨텍스트 영상을 CNN구조에 입력으로 넣는다. 첫 번째 컨볼루션 층은 stride 4에 $11 \times 11 \times 3$ 크기의 커널을 사용하여 96개의 특징 맵을 얻고, 두 번째 컨볼루션 층은 max-pooling을 통과한 $27 \times 27 \times 96$ 크기의 영상을 입력으로 사용하여 stride 2에 $3 \times 3 \times 96$ 크기의 커널을 통해 256개의 특징 맵을 얻는다. 두 번째 컨볼루션 층의 결과에 대해 다시 max-pooling을 시킨다. 컨볼루션 층에 이어져 있는 완전 연결 층은 각각 4096개와 1개의 뉴런으로 구성 된다. 마지막으로 인간의 인지된 안전도 점수를 예측하기 위해서 마지막 완전 연결 층의 뉴런 수를 1개로 설정하였고, 손실은 유클리디안 손실(Euclidean loss)을 통해 계산하였다.

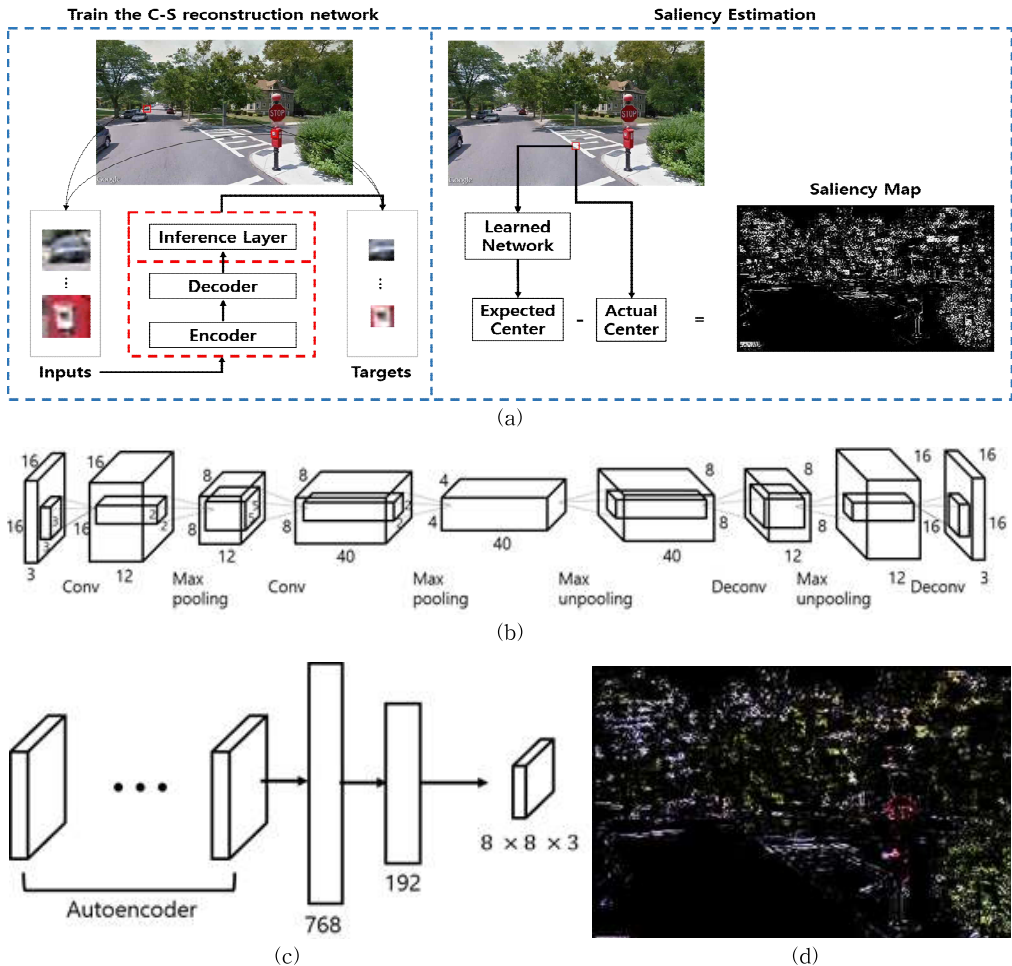


Fig. 3. (a) Diagram of Context extraction algorithm, It consists three parts to extract environmental context. (b) The structure of C-S reconstruction network, Using SCA, it is trained by inputs and corresponding targets, (c) After autoencoder, we change the size of last fully connected layer to predict center of inputs, (d) Using context map, we extract environment context in original image.

5. 실험 방법 및 결과

5.1 실험 방법

본 논문에서는 도시 영상에 대한 안전도 점수를 예측하기 위해서 컨텍스트를 반영한 CCNN 구조를 사용한다(Fig. 5). CCNN과 기존 연구에서 사용된 다양한 CNN모델들, 그리고 SVR 모델의 성능을 평균 제공근 오차(MSE)와 상관관계 분석을 바탕으로 비교하는 실험을 진행하였다. CNN 모델은 Alexnet과 특징 공유 구조(Feature sharing)를 사용하였다. 데이터 셋은 Place Pulse 1.0 데이터 셋과 우리가 수집한 데이터 셋으로부터 총 2,286장을 사용하였으며,

이 중 학습에 1,900장, 테스트에 386장을 사용하였다. SVR은 L2 loss를 사용하는 L2-regularized SVR 모델을 사용하였고, 파라미터 c 와 p 를 변경하면서 학습을 실시하였다. 각각의 CNN 모델에 대한 학습율은 0.001, dropout 비율은 0.7로 설정하였다. 학습을 위한 딥러닝 프레임워크로 Tensorflow[23]를 사용하였다.

5.2 실험 결과

우리는 여러 모델들 간의 성능 비교를 진행하기 위해서 평균 제공근 오차(MSE)와 상관관계 분석을 실시하였다. 평균 제공근 오차는 ground truth로 가

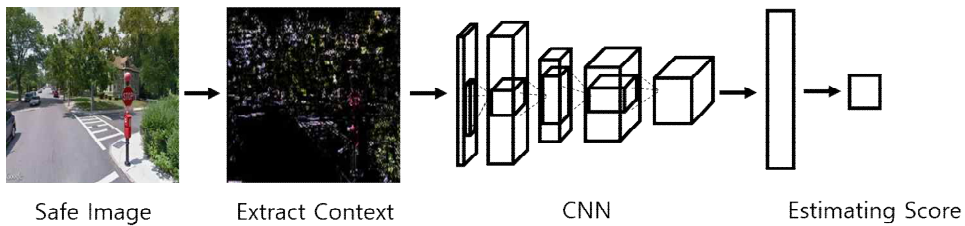


Fig. 4. The structure of CCNN case of training safe image. First, extract the context from input image. Second, Extract features using learned network, Finally, concatenate features and predict safety score.

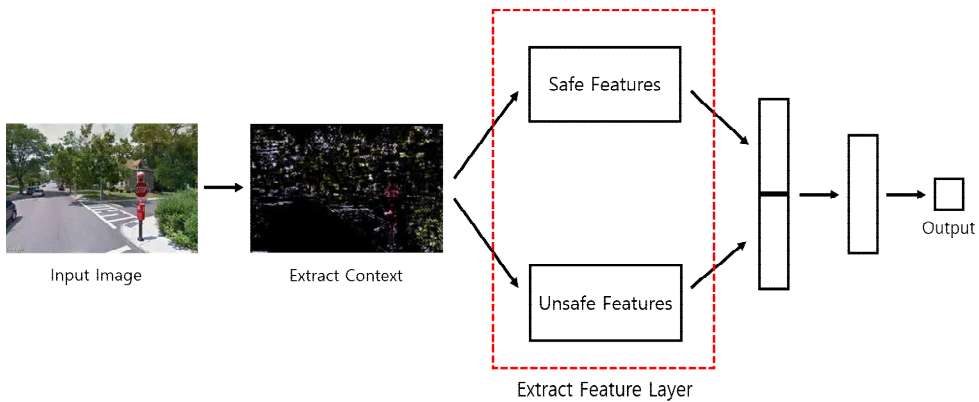


Fig. 5. The structure of CCNN. It consists of context, safe feature, unsafe feature and joint feature representations layers.

지고 있는 원 안전도 점수와 모델을 통해 얻은 예측된 안전도 점수의 차이가 얼마나 나는지를 측정하는 방법이다.

Table 1은 평균 제곱근 오차와 상관계수에 따른 성능 비교 결과이다. 비교 결과를 보면 우리의 CCNN 모델이 0.7548의 평균 제곱근 오차를 가지며 가장 좋은 성능을 보여주었다. 다시 말해서 우리가 제안하는 모델의 예측 결과가 다른 모델들에 비해 가장 적은 차이를 가짐을 알 수 있다. 또한 모델의 예측 안전도 값과 원 안전도 점수 사이의 상관관계 분석을 통해 CCNN 모델의 추가적인 성능 비교를 하였다. 상관관계 분석은 피어슨 상관계수(Pearson correlation co-

efficient)와 스피어맨 상관계수(Spearman correlation coefficient)를 통해 실시하였고, 통계 분석은 SPSS ver 22을 사용하였다. Fig. 6는 예측된 안전도 점수와 원 안전도 점수의 산점도를 보여준다. 모든 모델의 유의 확률이 0.01보다 작은 값을 가지는 것을 통해 유의수준 0.01에서 유의미하다는 것을 확인할 수 있다. CCNN 모델은 0.851/0.843의 피어슨/스피어맨 상관계수로 예측된 안전도 점수와 원 안전도 점수가 가장 높은 양의 상관관계를 가짐을 확인할 수 있고, 이를 통해서 다른 모델에 비해 정확하게 점수를 예측한 것을 볼 수 있다.

Table 1. Results of RMSE and correlation coefficient performance comparison

Model	RMSE	Pearson correlation coefficient (p-value)	Spearman correlation coefficient (p-value)
SVR	0.9287	0.746 (< 0.01)	0.751 (< 0.01)
Alexnet	0.8504	0.769 (< 0.01)	0.769 (< 0.01)
Sharing	0.7726	0.829 (< 0.01)	0.824 (< 0.01)
CCNN	0.7548	0.851 (< 0.01)	0.843 (< 0.01)

6. 결 론

본 논문에서는 도시 영상으로부터 인지도 안전도 점수를 예측하는 새로운 모델인 CCNN 구조를 제시하였다. 기존의 CNN 구조로는 추상적인 정보로부터 정확한 값을 예측하는데 어려움이 있는 문제가 있었는데, 우리는 영상으로부터 추출한 컨텍스트를 이용한 새로운 학습 및 예측 방법을 사용하였다. 기존 모델과의 성능 비교를 위해 SVR, Alexnet, Sharing 3 가지 모델을 학습하였고, 비교 결과 0.7548의 평균 제곱근 오차로 가장 좋은 성능을 보였다. 또한 예측 점수와 원 안전도 점수의 상관관계 분석을 통해 높은 양의 상관관계를 보임을 확인하였다.

컨텍스트를 추출하는 과정에서 입력으로 들어온 영상에 대해 Saliency map을 구할 때, 입력 영상의 컨텍스트가 뚜렷할 경우 안전도를 예측함에 있어 적은 오차를 보이며 좋은 결과를 얻을 수 있었다. 다만 컨텍스트가 뚜렷하지 않은 일부 경우에 있어 추출된 컨텍스트 영상의 픽셀 값이 0인 공간의 비중이 늘어질 경우, 이를 반영한 신경망의 성능을 저하시키는 것을 확인할 수 있었다. 따라서 컨텍스트를 추출하는 과정에서 추출된 컨텍스트 영상의 빈공간은 줄이고 판단의 핵심이 되는 부분의 신호는 강하게 나타내어 전체적인 성능을 향상시킬 수 있도록 개선되어야 한다.

범죄 기록, 설문지 조사가 아니라 도시 영상으로부터 안전도를 예측할 수 있게 함으로써 도시 안전도 예측에 필요한 시간적, 물질적 재원을 더욱 효율적으로 운영하는데 도움을 줄 수 있다. 또한 거리에 있던 많은 낙서를 지우고, 도시를 깨끗하게 만드는 일을 통해 강력 범죄 발생률을 줄인 뉴욕의 사례와 같이 ‘안전하지 않다’라고 판단하게끔 만드는 컨텍스트에 대한 분석을 통해 범죄예방 환경설계(CPTED: Crime Prevention Through Environmental Design)[24]을 진행하는데 큰 도움을 줄 것이다. 실험 데이터를 분석해본 결과 공사 중인 장소, 지저분한 거리, 텅 빈 거리 영상이 주로 낮은 안전도 점수를 가지고, 가로수가 많거나 주택가, 깔끔한 거리, 사람이 많은 영상이 주로 높은 안전도 점수를 가지고 있음을 확인할 수 있었다.

이후의 연구에서는 국내 도시에 대해 데이터를 추가적으로 수집할 계획이고, 이러한 데이터 셋을 이용해서 서울, 인천 등 국내 대도시를 대상으로 안전도 수치 맵을 구성할 계획이다. 또한 현재는 속도보다

정확도를 중시했는데, 좀 더 큰 데이터에 대해서 효율적으로 예측하기 위해서는 이미지 영상을 효율적으로 다루기 위한 시도[25]처럼 보다 빠르게 학습을 할 수 있으면서 정확도를 유지할 수 있는 네트워크의 개발이 필요하다.

REFERENCE

- [1] Chatterton, Paul, and R. Hollands, *Urban Nightscapes: Youth Cultures, Pleasure Spaces and Corporate Power*, Psychology Press, Hove, United Kingdom, 2003.
- [2] M. Roberts and A. Eldridge, *Planning the Night-time City*, Routledge, Abingdon, United Kingdom, 2012.
- [3] S.H. Kim, "The Impact of Local Festivals on Local Economy," *Korean Public Administration Review*, Vol. 39, No. 4, pp. 221-246, 2005.
- [4] G. Helms. *Towards Safe City Centres?: Remaking the Spaces of an Old-industrial City*, Ashgate Publishing Ltd, Aldershot, United Kingdom, 2012.
- [5] J. Bachner, *Predictive Policing: Preventing Crime with Data and Analytics*, IBM Center for the Business of Government, Washington, D.C., 2013.
- [6] The Santa Cruz Experiment: Can a City's Crime Be Predicted?(2011). <http://www.popsci.com/science/article/2011-10/santa-cruz-experiment> (accessed Jul., 16, 2017).
- [7] J.Q. Wilson and G.L. Kelling, "Broken Windows," *Atlantic Monthly*, Vol. 249, No. 3, pp. 29-38, 1982.
- [8] G.O. Mohler, M.B. Short, P.J. Brantingham, F.P. Schoengberg, and G.E. Tita. "Self-exciting Point Process Modeling of Crime," *Journal of the American Statistical Association*, Vol. 106, No. 493, pp. 100-108, 2011.
- [9] P. Salesses, K. Schechtner, and C.A. Hidalgo, "The Collaborative Image of the City: Mapping the Inequality of Urban Perception," *PloS one* Vol. 8, No. 7, e68400, 2013.
- [10] Ordonez, Vicente, and T.L. Berg. "Learning

- High-level Judgments of Urban Perception," *European Conference on Computer Vision*, pp. 494-510, 2014.
- [11] A. Khosla, B.A. An, J.J. Lim, and A. Torralba, "Looking Beyond the Visible Scene," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3710-3717, 2014.
- [12] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo, "Streetscore-predicting the Perceived Safety of One Million Streetscapes," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 779-785, 2014.
- [13] G.S. Matthew, "Predicting Crime Using Twitter and Kernel Density Estimation," *Decision Support Systems*, Vol. 61, pp. 115-125, 2014.
- [14] Chen, Xinyu, Y.W. Cho, and S.Y. Jang. "Crime prediction Using Twitter Sentiment and Weather," *Proceeding of Systems and Information Engineering Design Symposium*, pp. 63-68, 2015.
- [15] Xia, Chen, F. Qi, and G. Shi. "Bottom-up Visual Saliency Estimation with Deep Autoencoder-based Sparse Reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 27, No. 6, pp. 1227-1240, 2016.
- [16] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber, "Stacked Convolutional Auto-encoders for Hierarchical Feature Extraction," *Artificial Neural Networks and Machine Learning - ICANN 2011*, pp. 52-59, 2011.
- [17] S. Chen and T. Joachims, "Predicting Match-ups and Preferences in Context," *KDD*, pp. 775-784, 2016.
- [18] R. Herbrich, T. Minka, and T. Graepel, "TrueSkillTM: A Bayesian Skill Rating System," *Advances in Neural Information Processing Systems*, pp. 569-576, 2007.
- [19] O. Aude and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, Vol. 42, No. 3, pp. 145-175, 2001.
- [20] P. Florent, J. Sánchez, and T. Mensink, "Improving the Fisher Kernel for Large-scale Image Classification," *Computer Vision - ECCV 2010*, pp. 143-156, 2010.
- [21] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, et al. "Decaf: A Deep Convolutional Activation Feature for Generic Visual Recognition," *Proceeding of International Conference on Machine Learning*, pp. 647-655, 2014.
- [22] D. Navneet and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 886-893, 2005.
- [23] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al. "Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems," *arXiv Preprint arXiv:1603.04467*, 2016.
- [24] P.M. Conzens, G. Saville, and D. Hillier, "Crime Prevention through Environmental Design (CPTED): A Review and Modern Bibliography," *Property Management*, Vol. 23, No. 5, pp. 328-356, 2005.
- [25] I.H. Choi and S.H. Lee, "Content-based Image Retrieval using Reature Extraction in Wavelet Transform Domain," *Journal of Korea Multimedia Society*, Vol. 5, No. 4, pp.415-425, 2002.



이 기 인

2013년~현재 가톨릭대학교 미디어기술콘텐츠학과 학사과정

관심분야: 컴퓨터비전, 기계학습, 인공지능, 빅데이터



강 행 봉

1980년 한양대학교 전자공학과(학사)

1986년 한양대학교 전자공학과(석사)

1989년 Ohio State Univ. 컴퓨터공학(석사)

1993년 Rensselaer Polytechnic Institute 컴퓨터공학(박사)

1993년~1997년 삼성종합기술원 수석연구원

1997년~현재 가톨릭대학교 미디어기술콘텐츠학과 교수

2005년 UC Santa Barbara, Visiting Professor

관심분야: 컴퓨터비전, 기계학습, HCI, 컴퓨터그래픽스, 인공지능, 빅데이터