

자연화상에서 FFNet를 리용한 본문령역검출방법

주진혁, 리일홍

경애하는 김정은동지께서는 다음과 같이 말씀하시였다.

《첨단과학기술분야에서 세계적경쟁력을 가진 기술들을 개발하기 위한 투쟁을 힘있게 벌려야 합니다.》

지난 시기에는 자연화상에서 본문령역을 보통 본문령역후보검출단계와 본문령역검증단계의 두 단계를 거쳐 검출하였다. 본문령역후보검출단계[1]에서는 본문령역이 가지고있는 특징을 수학적으로 모형화하고 그것을 리용하여 본문이 들어있을수 있는 4각형령역들을 본문령역후보로 선택한다. 본문령역검증단계[2]에서는 후보로 선택된 4각형령역들에 대하여 본문령역인가 아닌가를 판정하고 본문령역이 아닌 후보들을 제거한다. 이와 같은 방법은 본문령역의 특징추출과 후보들에 대한 검증의 제한성으로 하여 본문령역검출에서 정확도가 높지 않았다.

본문에서는 선행방법들의 부족점을 극복하고 대조도가 낮은 화상과 잡음이 있는 화상들에서도 본문령역을 높은 정확도로 검출할수 있도록 심층신경망모형인 FFNet(Fast Forwarding Net)를 리용한 본문령역검출방법을 제안하였다.

1. FFNet를 리용한 본문검출모형

FFNet를 리용한 본문검출모형에서는 본문의 특징을 얻을수 있는 중첩층들을 리용하여 본문령역을 정확히 검출하면서도 학습속도를 개선한다.

1) FFNet의 개념

FFNet는 연속적인 단계를 형성하는 중첩층들의 모임이다.

하나의 빠른 정방향전파단계를 그림 1에 보여주었다. 그림 1에서 보여준것처럼 계산은 2개의 병렬경로를 통해 수행된다. 왼쪽가지는 표준적인 중첩층들로 이루어진 경로이며 오른쪽가지는 병렬자료경로이다.

마디 $S1$ 은 입력이고 마디 $S2$ 는 출력이다. 왼쪽가지는 중첩층, 오른쪽가지는 빠른 정방향전파가지(FF)이다. 중첩층은 크기가 $3 \times 3 \times 64$ 이며 매 층에서 활성화함수로 정규화선형신경세포(ReLU)를 리용한다.

매 단계에서 중첩층을 $S2C1$, $S2C2$, $S2C3$ 이라고 하면 $S2C3$ 은

$$S2C3 = \text{CONV}_{3 \times 3, s=1, p=1}(\text{CONV}_{3 \times 3, s=1, p=0}(\text{CONV}_{3 \times 3, s=1, p=1}(S1)))$$

과 같이 계산된다. 여기서 s 는 stride, p 는 자료저장층이 있는가 없는가를 의미한다.

$S2C3$ 의 크기는 $S1$ 이 $N \times N \times C$ 라고 하면 $(N-2) \times (N-2) \times 64$ 이다.

FF가지는 $5 \times 5 \times 64$ 중첩층과 ReLU활성화함수를 리용한다. $S1$ 을 입력으로 하고 $B2C1$ 을 출력으로 한다.

$$B2C1 = \text{CONV}_{5 \times 5, s=1, p=0}(S1)$$

중첩층가지와 FF가지는 하나의 출력 $S2$ 를 위해 연결된다. $B2C1$ 의 크기도

$$(N - 2) \times (N - 2) \times 64$$

이므로 $S2$ 의 크기는

$$(N - 2) \times (N - 2) \times 128$$

이다. 즉 낮은준위와 높은준위표현의 합이 다음층으로 넘어가게 된다.

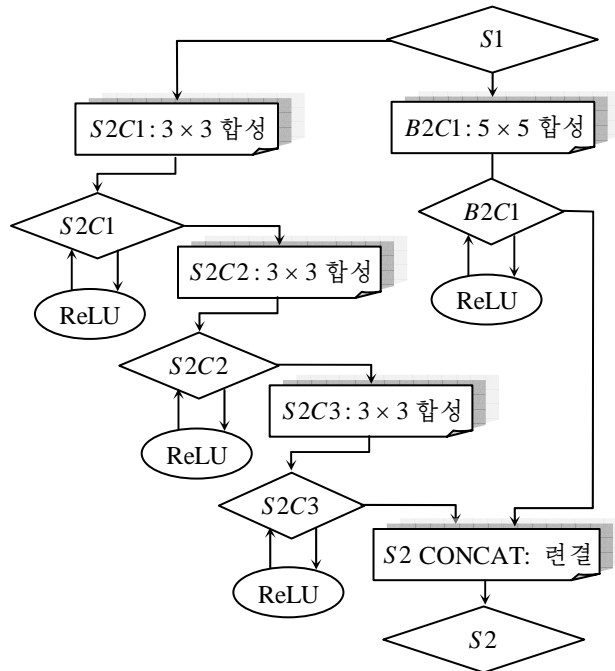


그림 1. 하나의 빠른 정방향전파단계

FFNet를 리용하여 본문검출을 진행하는 모형을 그림 2에 보여주었다.

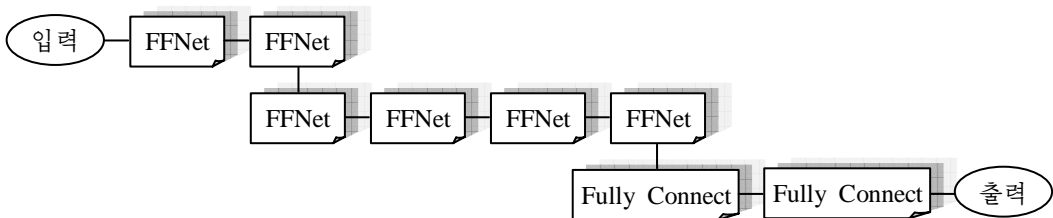


그림 2. FFNet를 리용하여 본문검출을 진행하는 모형

2) FFNet의 학습

학습단계 1 본문영역을 검출하기 위해 본문과 배경으로 이루어진 실험학습자료기지 (TDB: Test Database)를 리용한다.

본문영역검출을 위한 학습모형의 구조를 그림 3에 보여주었다.

학습자료기지는 일반배경에 본문들을 배치하여 컴퓨터로 생성한것과 사진화상자료들을 결합하여 구축한다.

학습단계 2 본문영역검출을 위해 학습단계 1에서 진행한 특징지도를 리용하여 마스크에 의한 학습을 진행한다.

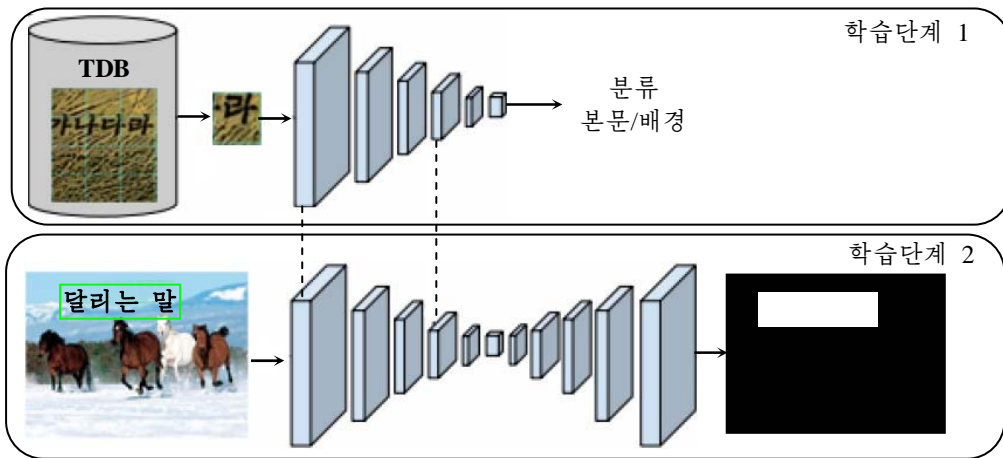


그림 3. 본문령역검출을 위한 학습모형의 구조

2. 실험 및 결과

본문검출에 리용하는 실험자료기지는 32pixel × 32pixel 크기의 본문화상과 배경화상들을 리용하여 만든다. 본문이 있는 령역은 본문령역, 본문이 없는 령역은 배경령역으로 한다. 실험자료기지는 학습자료기지 10만건, 검증자료기지 5만건, 시험자료기지 5만건으로 되어있다.

론문에서는 학습한 FFNet모형을 리용하여 본문의 마스크령역을 검출한다. 마스크령역은 본문을 포함하는 4각형령역을 의미하므로 그 령역을 검출하면 자연화상에서 본문령역을 구할수 있다.

종전의 방법으로 검출하기 어려운 일부 화상들에 대한 본문검출결과를 그림 4에 보여주었다.

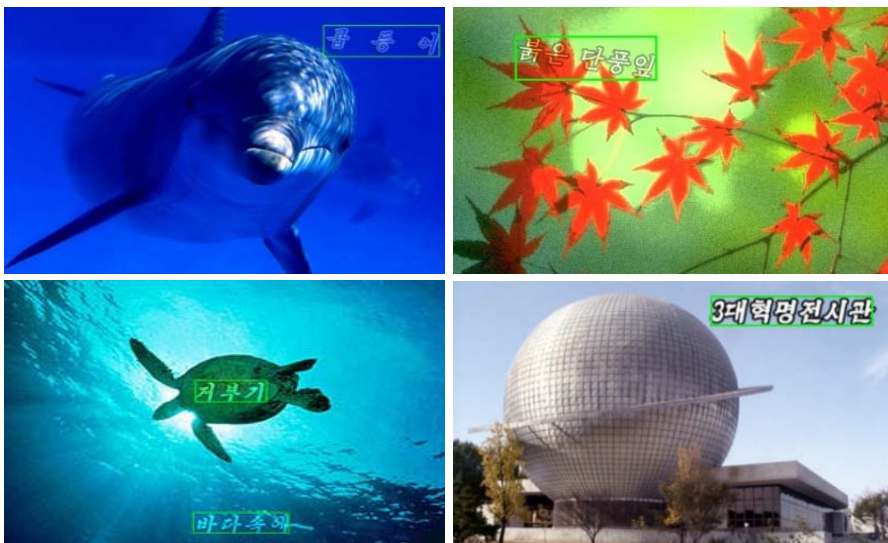


그림 4. 종전의 방법으로 검출하기 어려운 일부 화상들에 대한 본문검출결과

실험결과로부터 제안한 방법이 선행한 본문영역검출방법에서 제기되었던 대조도가 낮고 잡음이 있는 자연화상에서 본문영역검출문제를 해결할수 있다는것을 알수 있다.

GoogleLeNet와 CaffeNet와의 비교결과를 그림 5에 보여주었다.

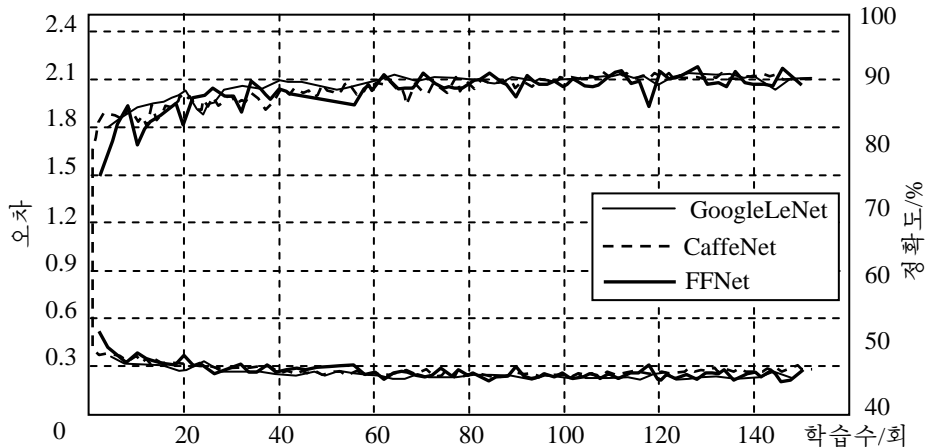


그림 5. GoogleLeNet와 CaffeNet와의 비교결과

맺 는 말

자연화상에서 본문영역을 검출하는 방법이 선행방법에 비하여 검출정확도가 상당히 높다는것을 실험을 통하여 확증하였다.

참 고 문 헌

- [1] Sezer Karaoglu et al.; IEEE Transactions on Multimedia, 19, 5, 1063, 2017.
- [2] Chong Yu et al.; IET Computer Vision, 9, 4, 603, 2014.

주체110(2021)년 5월 5일 원고접수

A Method for Text Detection Using FFNet on Natural Image

Ju Jin Hyok, Ri Il Hung

In this paper, a method of text detection was proposed using FFNet, Deep Learning Model and their performance was proved in experiment.

Keywords: text detection in natural image, deep learning, fast forwarding network