

객체 감지 신경망을 이용한 광학 점자 인식

Ilya G. Ovodov
 ELVEES RnD 센터, JSC 젤레노
 그라드, 러시아
 iovodov@elvees.com

추상적인

광학 점자 인식 방법은 일반적으로 점자 텍스트의 기하학적 구조에 크게 의존합니다. 이 구조가 왜곡되면 문제가 발생합니다. 따라서 스마트폰으로 촬영한 책 페이지 이미지에 대처하는 데 어려움을 겪는다.

객체 감지 컨볼루션 신경망을 이용하여 전체 점자 문자를 한 번에 감지하는 광학 점자 인식 방법을 제안합니다. 제안된 알고리즘은 이미지에 표시되는 점자 페이지의 변형 및 원근 왜곡에 강인합니다. 이 알고리즘은 국내 상황에서 스마트폰 카메라로 촬영한 점자 텍스트를 인식하는 데 적합합니다. 원근 왜곡이 있는 곡선 페이지와 이미지를 처리할 수 있습니다. 제안하는 알고리즘은 기존 방법에 비해 높은 성능과 정확도를 보인다.

또한 각 점자 문자에 대한 주석이 있는 점자 텍스트 사진 240장을 포함하는 새로운 데이터 세트를 생성했습니다. 제안된 알고리즘과 데이터 세트는 모두 GitHub에서 사용할 수 있습니다.

1. 소개

엠보싱 점자 알파벳은 1824년에 발명되었으며 수년 동안 시각 장애인을 위한 기본 쓰기 및 읽기 방법으로 사용되었습니다. 최근의 기술 발전은 시각 장애인에게 정보를 수신하고 전송할 수 있는 많은 새로운 기회를 제공합니다. 그럼에도 불구하고 인쇄된 점자를 읽고 점자로 쓰는 것은 여전히 그들에게 중요한 의사소통 방법입니다. 또한 점자는 시각 장애인과 시각 장애인 간의 의사 소통에 자주 사용됩니다.

특히 시각장애인 교사가 시각장애인 학생과 함께 일할 때 점자로 된 교과서와 학생 작품을 다루어야 하는 상황은 매우 흔하다.

점자는 촉각 인식을 위해 종이에 양각으로 점을 찍어 만든 알파벳입니다. 그래도 점자 텍스트를 다루는 정안인들은 대개 손가락으로 텍스트를 읽을 수 있는 촉각 능력이 부족하여 눈으로 읽어야 합니다.

시력이 좋은 사람들에게 점자 텍스트는 많은 작은 흰색 점자처럼 보입니다.



그림 1. 양면 인쇄로 찍은 페이지의 예.

흰색 배경에 튀어나온 점이 있어 시각적 인식이 매우 지루합니다. 양면 인쇄 점자를 읽는 것은 특히 어렵습니다. 페이지 뒷면에 인쇄된 텍스트는 움푹 패인 점처럼 보입니다. 촉각적으로 이 점들은 거의 보이지 않으며 전면 볼록한 점의 감각을 방해하지 않습니다. 그러나 육안으로는 거의 구별할 수 없습니다(그림 1 참조). 따라서 그러한 텍스트를 눈으로 읽는 것은 특히 어렵습니다. 기술적인 점자 인식 도구, 특히 광학 인식 방법을 사용하면 이 작업을 크게 용이하게 할 수 있습니다. 따라서 광학 점자 인식(OBR) 방법은 적어도 1980년대부터 발전해 왔습니다 [2].

각 문자 또는 기타 문자는 2x3 그리드에 여러(1~6) 돌출 점이 있는 점자 텍스트로 표시됩니다. 따라서 63개의 서로 다른 문자를 인코딩할 수 있습니다. 인접한 기호 사이의 거리는 하나의 기호에 있는 두 열의 점 사이보다 약간 더 깁니다. 문자 너비, 문자 사이의 간격, 줄 사이의 간격 및 줄에서 문자의 위치는 각 점자 문서에 대해 일정합니다. 따라서 점자 문자의 점은 고정 그리드의 노드에 위치합니다(그림 2). 이 기하학적 구조의 사용은 대부분의 기존 점자 인식 방법에 중요합니다. 그러나 이는 적용 가능성을 크게 제한합니다. 이러한 알고리즘은 점자 페이지의 정확한 정렬을 보장하기 위해 스캐너 또는 특수 촬영 조건이 필요합니다.

본 연구는 국내 환경에서 휴대폰 카메라로 획득한 점자 텍스트 이미지에 적용 가능한 인식 방법을 제공하는 것을 목표로 한다. 그리드 점자 문자

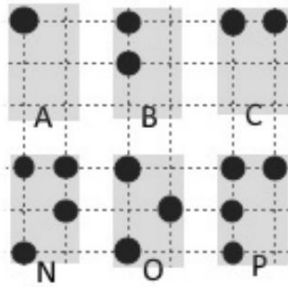


그림 2. 점자 텍스트의 도식적 배열.

a) 시트가 카메라 광축에 수직이 아니라는 사실로 인해 발생하는 원근 왜곡,
b) 펼쳐진 책 스프레드 이미지의 용지 곡률. 또한 시트의 다른 영역은 조명이 크게 다를 수 있습니다(그림 3).

이러한 문제를 해결하기 위해 객체 감지 합성곱 신경망 기반의 광학 점자 인식 알고리즘을 제안하였다. 제안된 알고리즘은 위에서 설명한 점자 페이지 이미지의 왜곡에 강합니다.

따라서 이 알고리즘은 일상적인 상황에서 스마트폰 카메라로 캡처한 점자 텍스트를 인식하는 데 적합합니다.
제안하는 알고리즘은 기존 방법에 비해 높은 성능과 정확도를 보였다.

또한 심층 신경망을 교육하고 인식 알고리즘의 정확도를 평가하는 데 사용할 수 있는 공개적으로 사용 가능한 데이터 세트가 부족하다는 문제에 직면했습니다. 사용 가능한 유일한 데이터 세트는 제한되어 있으며 위에서 설명한 경우를 포함하지 않습니다. 각 점자 문자에 대한 주석이 있는 점자 텍스트 사진 240장을 포함하는 새로운 데이터 세트를 생성했습니다.

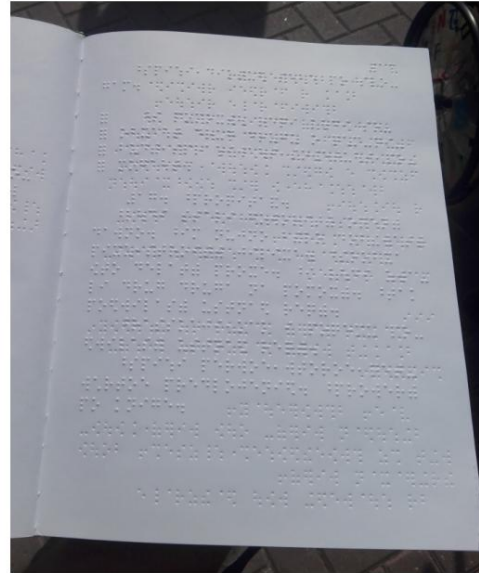
2. 관련업무

2.1. 광학 점자 인식 방식

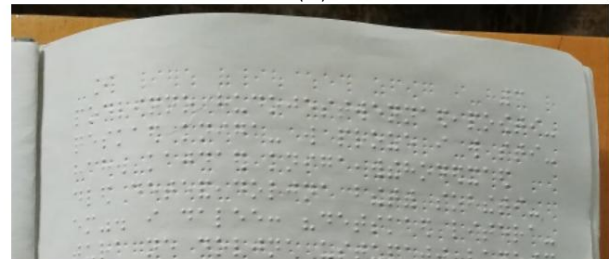
광학 점자 인식에 대한 주요 접근 방식은 점자 점 검색, 가능한 이미지 회전에 대한 보상을 포함하는 문자 그리드 복원, 문자로 점 그룹화, 마지막으로 문자 디코딩을 수행하는 순차적인 수행으로 구성됩니다. 조사 논문 [21, 5]은 이 일련의 단계를 따르는 모든 알고리즘을 고려합니다.

포인트 감지에 대한 가장 간단한 접근 방식은 임계값ing입니다. Zhang 과 Yoshino [24] 는 동적 로컬 임계값을 사용합니다. 양면 점자에서 점을 감지하고 앞면과 뒷면 점을 구별하기 위해 Antonacopou los와 Bridson [1] 은 점의 밝은 영역과 어두운 영역 감지를 사용하고 이미지를 밝은 영역, 어두운 영역, 배경 영역으로 구분합니다. . 포인트 주변의 평균 밝기 수준에 상대적인 정적 임계값을 사용합니다.

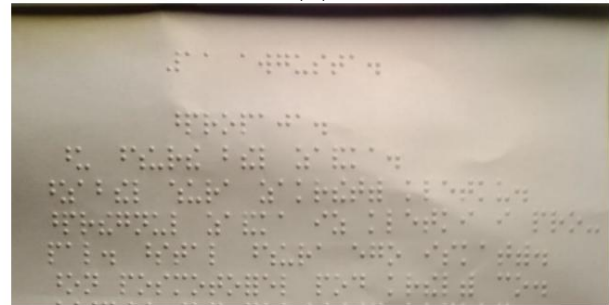
Morgavi와 Morando [16] 는 간단한 신경망을 사용하여



(a)



(b)



(c)

그림 3. 원근 왜곡(a), 페이지 곡률(b) 및 변형(c)이 있는 점자 텍스트 스마트폰 사진의 예.

포인트를 찾으십시오. Venugopal-Wairagade [22] 는 Hough 변환을 사용하여 원 감지를 수행합니다. Perera, Wanniarachchi [17] 는 HOG 및 SVM, R.Li et al. [10] - Haar 검출기 및 Adaboost., R.Li et al. [11] - 기본 도트 감지를 위한 Haar 감지기 및 Adaboost 및 그리드 복원 후 최종 도트 감지를 위한 HOG, LBP 및 SVM.

다음 단계는 포인트가 고정된 그리드를 복원하고 가능하면 시트 회전을 보상하는 것입니다.

이를 위해 선형 회귀 [17], 허프 변환 [1, 6] 또는 좌표 분포 밀도 기간 을 사용할 수 있습니다.

이미지의 단계별 회전을 조사하여 최적의 회전 각도를 찾을 수 있습니다 [11]. 때로는 이미지 기울기 보정 후 가능한 위치에 대한 정보를 사용하여 점을 다시 검색합니다 [1, 11].

일부 방법은 그리드 변형을 가정하여 서로 다른 선 사이의 피치 변화를 의미하지만 ([11] 및 기타) 그리드 선이 전체 시트에서 직선이고 평행하다고 가정합니다.

언급된 작업은 기본적으로 점자 점을 그리드에 스캔하는 데 의존하기 때문에 대부분 스캐너로 얻은 이미지로 작업합니다. 이 경우 필요한 이미지 보정이 기울기 보정으로 줄어들어 눈금선이 수직 및 수평이 됩니다. 소수의 작업만이 스마트폰의 이미지에 대한 OBR의 목적을 선언하지만 ([24, 22]) 이들이 설명하는 방법은 여전히 시트에 있는 직사각형 그리드의 존재를 기반으로 합니다.

CNN(컨볼루션 신경망)이 최근 몇 년 동안 이미지 인식 분야에서 엄청난 발전을 이루었지만 OBR에 딥 러닝 및 신경망을 사용하는 경우는 드뭅니다. 완전히 연결된 신경망의 사용에 대한 논문은 몇 개에 불과합니다. Morgavi, Morando [16] 및 Ting Li et al. [13] 간단한 신경망을 사용하여 포인트를 찾습니다. Subur et al. [20] - 영상 분할로 찾은 점들을 이용하여 심볼 값을 찾는다. Kawabe et al. [6] 양면 점자 인식 시 앞뒤 지점을 구분하기 위해 사용한다. R.Li et al. [12] CVPRW 2020에서 발표된 수정된 UNet 아키텍처를 사용한 분할 신경망 작업을 기반으로 합니다. 그들은 점자 문자가 차지하는 영역을 결정하고 이러한 문자를 인식하기 위해 신경망 작업을 사용했습니다. 분할 결과를 기반으로 개별 문자의 위치를 결정하려면 후속 후 처리가 필요합니다.

2.2. 광학식 점자 인식 방식 정확도 평가

다양한 연구에서 제안된 알고리즘의 정량적 정확도 특성이 서로 다르며 때로는 매우 높지만 알고리즘을 서로 비교하면 적어도 두 가지 장애물에 직면합니다.

- 최근까지 서로 다른 알고리즘을 비교할 공개 데이터 세트가 없었습니다. 논문에 제공된 품질 값은 공개된 데이터 세트가 아닌 독점 데이터에서 측정되었으므로 서로 다른 작업의 공개된 결과를 비교하는 것은 대부분 불가능합니다.
- 위에서 설명한 일반적인 파이프라인을 사용하는 작업(즉, 포인트 감지-그리드 복원-포인트를 기호로 그룹화 및 디코딩)을 사용하는 경우가 많으며 품질 지표는 포인트 인식 단계에만 제공됩니다. 우리 알고리즘과 같이 별도의 포인트 인식 단계가 없는 알고리즘과 이러한 알고리즘의 성능을 비교하는 것을 방지합니다.

점자 텍스트가 포함된 공개적으로 사용 가능한 유일한 DSBi 데이터 세트는 Li et al.에 의해 게시되었습니다. [10]. 기차(26페이지)와 테스트(88페이지) 세트로 나누어진 114페이지의 스캔 양면 점자 텍스트가 포함되어 있습니다. 스캔하는 동안 모든 페이지가 조심스럽게 정렬됩니다. 앞면과 뒷면에 대한 포인트 그리드가 계산되었으며 그리드를 수직-수평 방향으로 가져오는 데 필요한 회전이 계산되었습니다.

주석은 회전 각도, 회전 후 수직 및 수평 그리드 선의 좌표, 이 그리드의 노드를 참조하는 점자 문자 목록을 지정하여 작성됩니다. 모든 텍스트는 중국어로 되어 있지만 점자 알파벳은 모든 언어에서 동일한 구조를 가지며 이 데이터 세트는 언어에 관계없이 사용할 수 있습니다.

이 데이터 세트는 인식 알고리즘의 전체 교육을 제공할 만큼 충분히 크고 가변적이지 않지만(교육 세트의 26페이지만) 문제에 대한 다양한 접근 방식을 비교할 수 있습니다. [10]의 저자는 이미지 분할(Antonacopoulos et al., [1]), Haar 기능과 Adaboost(Viola & Jones [23]), 알고리즘(Li et al. [23])에 기반한 알고리즘의 정확도를 비교했습니다. [11]). 그러나 포인트 감지의 정확도만 제공합니다. [12]에서 그들은 도트 수준이 아니라 문자 수준에서 추정되는 알고리즘의 정확도 메트릭을 제공했습니다.

2.3. 객체 감지 컨볼루션 신경망

컴퓨터 비전, 특히 객체 감지에서 CNN(컨볼루션 신경망)의 사용은 최근 몇 년 동안 엄청난 발전을 이루었습니다. 컨볼루션 네트워크는 1989년 LeCun[9]에 의해 제안 되었지만 2012년부터 폭발적인 인기를 얻었습니다 [8]. 분류 문제 이후, 객체 감지 문제를 해결하기 위해 적용되었습니다. 즉, 이미지에서 객체를 포함하는 직사각형 영역을 동시에 찾고 그 안에 포함된 객체를 분류하는 것입니다. 처음에 개체 감지를 위한 CNN 기반 솔루션은 영역 검색과 개체 분류를 별도로 처리했습니다 ([4] 및 기타). 지역과 그 분류를 동시에 검색하는 이후의 방법(1단계 검출기)은 큰 성공을 거두었습니다.

객체 감지를 위한 1단계 감지기 원리는 SSD ([15, 3]) 및 YOLO ([18, 19]) 컨볼루션 신경망 아키텍처의 출현으로 널리 보급되었습니다. 1단계 감지기의 핵심 아이디어는 일련의 컨볼루션 및 크기 축소 후에 가능 맵을 생성한다는 것입니다. 이 지도의 각 점은 원본 이미지의 직사각형 영역에 해당합니다. 학습하기 전에 앵커라고 하는 가능 맵의 각 셀에 대해 원하는 개체의 여러 선택적 크기가 설정됩니다. 피쳐 맵의 각 포인트(즉, 원본 이미지의 각 직사각형 영역)에 대해 앵커와 교차하는 ground truth 주석 상자를 고려하고 교차 정도를 IOU(intersection over union)로 계산합니다. 신경망은 앵커의 필요한 이동 및 크기 조정을 예측하는 방법을 학습합니다(4개 중

앵커당 넣기), 교차 정도(앵커당 출력 1개), 앵커 영역에 개체의 존재를 나타내는 추론 단계에서 신뢰도 측정과 관련되며 개체의 클래스(여기서 C 출력 매개변수 C는 클래스 수입니다.) 따라서 가능 맵의 각 지점에서 신경망은 $A \cdot (4 + 1 + C)$ 출력 값을 예측하는 방법을 학습합니다. 여기서 A는 앵커 수이고 4개의 출력은 경계 상자의 좌표를 정의합니다 ([18, 19]) 또는 $A \cdot (4 + C)$ 출력 값 ([15, 3, 14]). 후자의 경우, 모든 C 클래스에 대한 응답에 신뢰도가 포함되어 영역이 개체에 해당하지 않는 앵커의 경우 모든 C 클래스에 대한 응답이 작습니다.

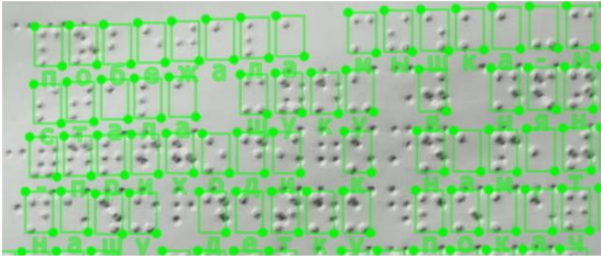


그림 4. Angelina Braille 이미지 데이터 세트의 샘플 주석.

손실 함수를 사용하여 위치 및 클래스 예측의 동시 학습이 이루어집니다.

(1)

여기서 L_{loc} 은 위치 예측 오류에 대한 손실 함수, L_{cls} 는 분류 오류에 대한 손실 함수, λ_{cls} 는 가중 계수입니다.

RetinaNet [14]은 더 복잡한 경우에 더 많은 가중치를 제공하는 손실 함수의 L_{cls} 구성 요소에 대해 개선된 손실 함수 FocalLoss를 제안하는 이 접근 방식을 추가로 개발합니다. 자세한 내용은 [14]를 참조하십시오. 우리는 이 작업에서 객체 감지 CNN 구현을 사용했습니다.

3. 우리의 접근 방식

3.1. 문제 설정과 네트워크 아키텍처

기존의 접근 방식과 달리 도트 감지, 그리드 복원 및 도트를 문자로 결합하는 단계를 분리하지 않습니다. 대신 위에서 설명한 개체 감지 CNN을 사용하여 전체 점자 기호를 직접 찾고 동시에 인식합니다. 공식 $c = 2^i - 1$ 을 사용하여 1에서 63까지의 클래스 레이블을 각 문자에 할당합니다.

그렇지 않으면 0입니다.

입력 이미지를 100dpi 해상도로 조정합니다. 따라서 표준 A4 페이지는 약 864x1150 해상도로 조정됩니다. 점자문자는 가로간격 약 25pt, 줄간격 약 40pt로 배치한다.

우리는 [14]에 설명된 RetinaNet CNN 아키텍처를 약간 수정하여 사용합니다. 광학 점자 인식 작업은 너비 대 높이 비율이 고정된 거의 동일한 작은 크기의 물체를 다수 검색한다는 점에서 다릅니다. 따라서 우리는 실행 시간, 주로 NMS 작업을 줄이기 위해 RetinaNet 아키텍처를 단순화했습니다. 16x16 크기의 가능 맵 셀이 있는 계층 수준에서 하나의 "클래스 + 상자 서브넷에 대한 출력"([14]의 그림 3 참조)만 사용되었습니다. 모든 점자 문자가 적어도 하나의 그리드 셀로 덮여 있음을 보장합니다. 크기가 가까운 각 그리드 셀에 대해 하나의 앵커만 사용했습니다.

예상 문자 크기로, 이러한 수정으로 인식 품질의 실질적인 손실 없이 계산 시간이 5배 이상 단축되었습니다.

또한 우리는 점자 문자가 겹치지 않고 NMS 절차를 사용하여 겹치는 감지된 경계 상자를 필터링하는 데 사용되는 IOU 임계값을 크게 줄인다는 우선적인 문제를 사용했습니다. 우리는 IOU 임계값 = 0.02를 사용하여 검출기 불완전성으로 인해 작은 겹침만 허용했습니다.

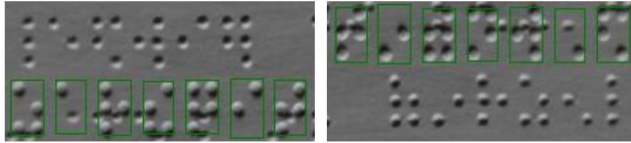
3.2. Angelina Braille 이미지 데이터 세트

DSBI 데이터 세트 [10]는 레이블이 지정된 점자 텍스트 이미지가 있는 공개적으로 사용 가능한 유일한 데이터 세트이며 작성자에게 깊은 감사를 표하고 싶습니다. 그러나 DSBI에는 점자가 직사각형 그리드에 정렬된 스캔된 점자 이미지만 포함됩니다. 다양성이 제한된 다소 적은 수의 이미지가 있습니다. 따라서 이 데이터 세트는 우리가 처리하고자 하는 어려운 이미지를 제대로 인식할 수 있는 CNN을 교육하는 데 적합하지 않으며, 이러한 어려운 샘플을 처리하도록 설계된 CNN을 테스트하는 데도 적합하지 않습니다.

212페이지의 양면 점자책과 28페이지의 학생 논문이 포함된 새로운 "Angelina Braille Images Dataset"을 준비했습니다. 이러한 텍스트는 알고리즘의 의도된 작업 조건에 가까운 조건에서 다양한 사진 카메라 또는 휴대폰으로 촬영되었습니다. 책의 펼친 부분에 구부러진 페이지와 원근 왜곡이 포함되어 있습니다. 텍스트의 앞면에 있는 문자는 일반적인 객체 감지 문제 방법을 사용하여 레이블이 지정됩니다. 각 문자에 대해 경계 상자가 정의되고 상자 내부의 점자 문자에 해당하는 1에서 63까지의 클래스가 할당됩니다. 샘플 이미지는 그림 3에, 샘플 라벨은 그림 4에 나와 있습니다.

데이터 세트의 주석이 반복적으로 생성되었습니다. 각 반복에서:

- 1. 기본 주석이 자동으로 생성되었습니다.
- 2. 점자 텍스트 라인에 맞게 자동으로 생성된 주석 경계 상자가 수동으로 수정되었습니다.
- 3. 주석 점자 문자가 일반 텍스트;



(가) (나)

그림 5. 이미지(a)가 180° 회전된 경우(b) 전면 돌출 점(녹색 직사각형 내부)은 후면 돌출 점처럼 보이고 그 반대도 마찬가지입니다.

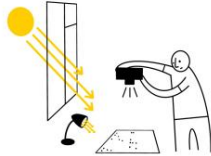


그림 6. 조명 조건 Angelina Braille 이미지 데이터 세트 이미지를 촬영했습니다.

4. 시 텍스트의 경우 텍스트를 줄로 나누는 것이 고정되어 있는 점을 고려하여 인터넷에서 찾은 동일한 시의 ground truth 텍스트와 비교하여 시 텍스트를 확인했습니다. 이후 반복에서 시가 아닌 텍스트는 맞춤법 검사기로 확인되었습니다. 질문 가능한 모든 사례를 수동으로 확인했습니다. 그 결과 어떤 위치에 점자가 있는지 없는지 명확하지 않은 경우 문법적으로 올바른 옵션을 선택하여 모호한 경우를 모두 해결했습니다.

우리는 돌출된 전면 점과 돌출된 후면 점은 떨어지는 빛의 방향을 알고 있을 때만 구별할 수 있음을 확인했습니다. 그림 5는 이미지가 180°회전된 경우 앞면 점은 뒷면 점처럼 보이고 그 반대도 마찬가지임을 보여줍니다. 페이지의 대략 상단 또는 상단 왼쪽에서 떨어지는 빛의 모든 이미지를 촬영하여 이러한 모호성을 피했습니다(그림 6).

데이터 세트는 191개의 훈련(80%) 이미지와 49개의 테스트(20%) 이미지로 나뉩니다. 또한 인터넷에서 가져온 다양한 비점자 텍스트 이미지 44개가 부정적인 예로 훈련 세트에 추가되었습니다.

점자 문자 클래스는 해당 러시아어 일반 텍스트 문자 및 기호로 표시됩니다. 그러나 이러한 레이블은 데이터 세트와 함께 제공되는 소프트웨어 도구를 사용하여 점자점으로 변환할 수 있습니다.

Angelina Braille 이미지 데이터 세트는 다음에서 사용할 수 있습니다.
GitHub: <https://github.com/IlyaOvodov/AngelinaDataset>.

4. 실험 설정 및 결과

4.1. 사용된 측정항목

문자 및 픽셀 수준에서 정밀도, 재현율 및 F1 메트릭을 평가했습니다. 알고리즘이 즉시 de

tects 기호, 기호 수준에서 정밀도를 평가하는 것은 당연합니다. 우리는 [12]와 일치하는 계산 방법을 사용했습니다. $IOU(intersection\ over\ union) \geq 0.5$ 인 실측 문자와 교차하고 올바른 클래스를 갖는 감지된 점자 문자는 참양성(TP)으로 간주됩니다. 그렇지 않으면 위양성(FP)으로 간주됩니다. 따라서 FP 감지에는 잘못된 위치 또는 올바른 위치가 있지만 잘못된 점자 문자가 할당된 감지가 모두 포함됩니다. 해당 TP 감지 점자 기호가 없는 Ground Truth 기호는 FN(거짓 부정)으로 간주됩니다. 정밀도, 재현율 및 F1 메트릭은 다음과 같이 정의됩니다.

우리는 또한 우리의 결과를 점 감지를 기반으로 하는 알려진 점자 인식 방법과 비교하기 위해 점당 메트릭을 평가했습니다. 알고리즘은 개별 포인트를 감지하지 않기 때문에 간접적으로 감지합니다. TP 캐릭터의 모든 포인트는 TP 포인트로 간주됩니다. $IOU \geq 0.5$ 로 검출이 교차하는 hot을 수행하는 실측 문자의 모든 지점은 FN 포인트로 간주되고 $IOU \geq 0.5$ 로 실측과 교차하지 않는 검출 문자의 모든 지점은 FP 포인트로 간주됩니다. $IOU \geq 0.5$ 인 실측 문자와 교차하지만 실측 문자 이외의 문자가 있는 탐지에 대해 gt 및 탐지된 점자 문자의 6개 위치 각각에서 점의 존재를 비교합니다. 두 문자 모두에 어떤 포인트가 있으면 TP로 간주됩니다. 그렇지 않으면 각각 FP와 FN으로 간주됩니다. 그런 다음 위에서 설명한 방식으로 도트 수준에서 Precision, Recall 및 F1을 평가합니다 (2).

4.2. 네트워크 교육

DSBI 데이터 세트 [10]를 사용하여 알고리즘의 효율성을 이전에 발표된 접근 방식 [10, 11, 12]과 비교했습니다. DSBI 데이터 세트에 대해 정의된 훈련-테스트 분할에는 28개의 훈련 이미지만 포함되며 이는 CNN 훈련에 너무 작습니다. 우리는 DSBI 데이터 세트에서 각 점자 책의 처음 74%로 훈련 세트를 정의하고 마지막 26% 페이지로 테스트 세트를 정의했습니다. 그 결과 열차 및 테스트 세트에 대해 각각 84개 및 30개의 이미지가 생성되었습니다.

보다 복잡한 조건에서 알고리즘을 평가하기 위해 위에서 설명한 DSBI 데이터 세트의 기타 세트와 Angelina Braille 이미지 데이터 세트를 결합했습니다. 평가는 DSBI 및 Angelina Braille Images Dataset 테스트 세트에서 개별적으로 수행되었습니다.

약 100dpi에 해당하는 864픽셀 너비로 크기 조정된 이미지를 처리하도록 신경망을 훈련했습니다. 신경망을 훈련할 때 다양한 이미지 스케일과 가능한 입력 왜곡에 대한 더 나은 저항을 얻기 위해 훈련 이미지를 다음과 같이 보강했습니다. 각

방법	학습 데이터 세트	테스트 데이터 세트	점자 도트 레벨				캐릭터 레벨				공연., s/이미지		
			정밀도	리콜	F1		Prec. 상가하다		F1				
Segment [10]	DSBI [10]	DSBI 0.9172	0.9811	0.948	0.9765	0.9638	0.970	0.9838					
머리카락 [10]					0.9575	0.970	0.9314	0.9869					
머리카락 [11]					0.958	0.9931	0.9997	0.996			0.89		
HOG,SVM [11]					0.9965	0.9997	0.998	0.9928			15.02		
SVM 그리드 [11]					0.9996	0.9962	0.9943	0.9988			1.22		
TS-OBR [11]					0.9966	DSBI 0.9992	0.9995	0.9994	0.9977	0.9975	0.9976	0.9984	1.45
브라넷 [12]									0.9993	0.9989	0.9961	0.9964	0.25
우리의	DSBI	0.9963	Our	0.9812	0.9143	0.9466	0.9569	0.8980	0.9265	DSBI+당사	당사	0.9995	0.18
	DSBI+우리			0.9986	0.9991	0.9985	0.9978	0.9981					
	DSBI												

표 1. 실험 결과. 점 및 문자 수준에서 테스트 데이터 세트의 세차, 재현율, F1 메트릭 및 처리 성능.

이미지는 필요한 너비의 ±30%인 550에서 1150pix 사이의 임의의 너비로 조정되었습니다. 그런 다음 이미지를 ±10% 이내의 무작위 배율로 세로로 압축하거나 늘렸습니다. 그런 다음 ±5° 내에서 임의의 각도로 이미지를 회전했습니다. 50%의 확률로 세로축을 따라 이미지를 반사하고 각 문자 레이블을 반사된 문자 레이블로 변경했습니다.

공식을 사용하여 이미지를 정규화했습니다.

(삼)

여기서 I_c 는 범위 [0,255]에 있는 이미지 채널의 강도이고, c, m 은 전체 이미지에 대한 I_c 의 평균이고 s 는 이미지에 대한 I_c 의 표준 편차입니다 .

무작위 416x416 이미지 자르기가 CNN 입력으로 사용되었습니다. 학습률 = 1e-4 및 배치 크기 = 24인 Adam 옵티마이저 [7] 를 사용하여 500 epoch 동안 신경망을 훈련했습니다 . 처음에는 손실 함수 (1) 의 λ_{cls} 계수를 1 로 설정했습니다. 이 경우 손실 함수의 Lloc 구성 요소가 Lcls 보다 우세하여 문자 분류보다 문자 위치 학습이 더 빠릅니다. 500 에포크 후에 우리는 λ_{cls} 를 100으로 설정하여 두 구성 요소의 기여도를 거의 비슷하게 만듭니다. 손실 함수의 두 구성 요소의 기여도가 처음부터 동일하게 설정되면 학습 프로세스가 불안정해집니다. 마지막으로 λ_{cls} = 1000으로 설정하고 "Reduce On Plateau" 접근 방식을 사용하여 CNN을 훈련합니다.

탐지 또는 실측 레이블이 올바른지 여부는 상당 부분 주관적입니다. 따라서 DSBi 데이터 세트에서 $F1 \geq 0.997$ 인 알고리즘의 추가 비교는 유익하지 않은 것 같습니다. 그 이유는 문자의 의미론적 의미를 고려하지 않고 DSBi 데이터 세트에 레이블을 지정했기 때문입니다. 따라서 모호한 사례는 임의로 분류되었습니다. 본 논문에서 제안하는 안젤리나 데이터셋은 점자 문자의 의미론적 의미를 이용하여 라벨링 하였으므로 문법적 관점에서 특정 위치에 존재해야 하는 옵션을 선택하여 의심스러운 경우를 해결하였다.

DSBi와 Angelina Braille Image Dataset 모두에서 교육을 받았을 때 우리의 방법은 An gelina 테스트 세트에서 $F1=0.9981$, DSBi에서 $F1=0.9963$ 이 됩니다. 이 상대적으로 동질적인 데이터 세트에서 일부 정확도 감소는 CNN이 더 다양한 데이터에 대해 훈련되어 더 많은 일반화 능력이 필요하기 때문이라고 가정할 수 있습니다. 보시다시피 이 데이터 세트의 사진에 나타나는 종이 변형 및 원근 왜곡은 페이지가 왜곡 없이 평평해지고 스캔되는 DSBi 데이터 세트보다 인식 품질을 악화시키지 않습니다.

제안된 방법으로 A4 페이지 1장을 처리하는 데 GPU NVIDIA 1080T에서 이미지당 0.18초의 시간이 소요됩니다. 동일한 하드웨어에 대해 0.25s/이미지가 보고되는 BraUNet [12] 을 포함한 다른 방법보다 성능 이 뛰어납니다.

우리 알고리즘의 소스 코드와 훈련된 네트워크 가중치는 GitHub: <https://github.com/IlyaOvodov/AngelinaReader>에서 확인할 수 있습니다.

4.3. 결과 및 토론

표 1 은 실험 결과를 보여준다.

DSBi 데이터 세트에서 교육을 받을 때 새 방법은 문자 기반 테스트에서 $F1=0.9976$ 을 제공하고 점 기반 테스트에서 $F1=0.9994$ 를 제공합니다. 다른 방법보다 성능이 뛰어납니다. 오류를 일으키는 문자를 조사하면 올바른 레이블이 종종 의심스러운 것으로 나타납니다(그림 7). 데시

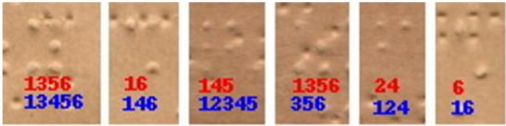


그림 7. 오류로 계산되는 DSBi 데이터 세트의 일부 문자. 빨간색 숫자 - 실제 점자 문자 점, 파란색 - 감지된 점자 문자 점. 어느 것이 맞는지는 명확하지 않습니다.

점자 텍스트 이미지 인식을 위한 공용 웹 서비스 Angelina Braille Reader를 출시했습니다. <https://angelina-reader.com>에서 볼 수 있습니다.

결론

본 논문에서는 객체 감지 합성곱 신경망 기반의 새로운 광학 점자 인식 알고리즘을 제안하였다. 제안한 알고리즘은 높은 정확도와 성능을 보였다. 이 알고리즘은 점자 텍스트가 있는 묘사된 시트의 불규칙성과 원근 왜곡에 저항하는 것으로 입증되었습니다. 따라서 일상적인 가정 환경에서 휴대폰으로 캡처한 텍스트를 인식할 수 있습니다.

이것은 우리의 알고리즘을 기존 알고리즘과 크게 구별합니다. 다른 방법은 점자 텍스트의 기하학적 구조가 호트러지지 않은 이미지가 필요합니다. 우리 알고리즘의 견고함은 눈으로 점자 텍스트를 읽어야 하는 사람들의 일상적인 요구에 사용할 수 있는 소프트웨어 서비스의 기반으로 사용할 수 있게 합니다. 맹인 학생의 교사, 시각 장애 아동의 부모 등.

주석이 달린 점자 텍스트가 있는 유일한 데이터 세트에는 위에서 설명한 목적을 위한 점자 인식 알고리즘의 성능을 평가하기에 어려운 샘플이 충분히 포함되어 있지 않습니다. 우리의 알고리즘과 앞에서 설명한 최상의 알고리즘은 모두 이 데이터 세트에 대해 거의 100% 정확도를 가집니다. 오류율은 모호한 주석 문자의 양과 비슷합니다.

새로운 Angelina Braille 이미지 데이터 세트를 만들고 게시했습니다. 스마트폰과 핸드 카메라로 찍은 더 어려운 샘플이 포함되어 있습니다. 데이터 세트에는 구부러진 책 스프레드 페이지, 원근 왜곡 이미지 및 기타 하드 케이스가 포함됩니다. 우리의 알고리즘은 이러한 더 어려운 이미지에서도 좋은 성능을 보입니다.

제안한 알고리즘인 새로운 Angelina Braille Image Dataset과 우리 알고리즘을 기반으로 한 광학 점자 인식을 위한 웹 서비스 Angelina Reader는 인터넷에서 사용 가능하다.

참조

[1] Apostolos Antonacopoulos와 David Bridson. 강력한 점자 인식 시스템. 문서 분석 시스템에 관한 국제 워크샵, 533–545페이지. 스프링거, 2004. 2, 3

[2] JP Dubus, M Benjelloun, V Devlaminck, F Wauquier 및 P Altmayer. 부조 점자를 검정 잉크로 변환하는 자동 시스템을 수행하는 이미지 처리 기술: Lectobracille. IEEE Engineering in Medicine and Biology Society의 연례 국제 회의 절차, 1584-1585 페이지. IEEE, 1988. 1 [3] Cheng-Yang Fu, Wei Liu, Ananth Ranga, Ambrish Tyagi 및 Alexander C Berg. Dssd: Deconvolutional 싱글 샷 검출기. arXiv 사전 인쇄 arXiv:1701.06659, 2017. 3, 4 [4] Ross Girshick, Jeff Donahue, Trevor Darrell 및 Jitendra Malik. 정확한 객체 감지를 위한 풍부한 기능 계층

그리고 사맨틱 세그멘테이션. 컴퓨터 비전 및 패턴 인식에 관한 IEEE 회의 진행, 페이지 580–587, 2014. 3 [5] Samer Isayed 및 Radwan Tahboub. 광학 점자 인식에 대한 검토. 2015년 웹 애플리케이션 및 네트워킹(WSWAN)에 관한 2차 세계 심포지엄, 1–6페이지. IEEE, 2015. 2

[6] 카와베 히로유키, 시모무라 유코, 난보 히데타카, 세토 슈이치. 오래된 점자책 복원을 위한 점자 분류에 답라닝 적용 경영 과학 및 엔지니어링 관리에 관한 국제 회의, 913-926페이지. Springer, 2018. 2, 3 [7] Diederik P Kingma 및 Jimmy Ba. Adam: 효율적 최적화 방법입니다. arXiv 사전 인쇄 arXiv:1412.6980, 2014. 6

[8] Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton. 깊은 컨벌루션 신경망을 사용한 Imagenet 분류 작업. 신경 정보 처리 시스템의 발전, 25:1097–1105, 2012. 3 [9] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard 및 Lawrence D Jackel. 손글씨 10개의 우편번호 인식에 적용된 역전파. 신경 계산, 1(4):541–551, 1989. 3

[10] Renqiang Li, Hong Liu, Xiangdong Wang, Yueliang Qian. Dsbi: 점자 점 감지를 위한 양면 점자 이미지 데이터 세트 및 알고리즘 평가. 2018년 비디오 및 이미지 처리에 관한 제2회 국제 회의 진행, 65–69페이지, 2018. 2, 3, 4, 5, 6

[11] Renqiang Li, Hong Liu, Xiangdong Wang, Yueliang Qian. 양면 점자 이미지에 대한 2단계 학습 기반의 효과적인 광학 점자 인식 Pacific Rim International Conference on Artificial Intelligence, 150-163페이지. 스프링거, 2019. 2, 3, 5, 6

[12] Renqiang Li, Hong Liu, Xiangdong Wang, Jianxing Xu, Yueliang Qian. 보조 학습 전략을 사용하는 사맨틱 분할 네트워크를 기반으로 한 광학 점자 인식.

컴퓨터 비전 및 패턴 인식 워크숍에 관한 IEEE/CVF 컨퍼런스 절차, 554–555페이지, 2020. 3, 5, 6

[13] Ting Li, Xiaoqin Zeng 및 Shoujing Xu. 점자 인식을 위한 딥 러닝 방법. In 2014 International Conference on Computational Intelligence and Communication Networks, 페이지 1092–1095, 2014. 3 [14] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollar. 조밀한 물체 감지를 위한 초점 손실. 컴퓨터 비전에 관한 IEEE 국제 회의의 시딩에서, 페이지 2980–2988, 2017. 4 찬성

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C Berg. Ssd: 단일 샷 멀티박스 감지기. 컴퓨터 비전에 관한 유럽 회의에서 21-37페이지. 스프링거, 2016. 3, 4

[16] 지오반나 모르가비와 마우로 모란도. 광학 점자 인식을 위한 신경망 작업 하이브리드 모델. 2002년 신호, 음성 및 이미지 처리에 관한 국제 회의(ICOSSIP 2002). 인용자, 2002. 2, 3

[17] TDSH Perera 및 WKIL Wanniarachchi. 지향성 기울기 특징 및 서포트 벡터 머신의 히스토그램을 기반으로 한 광학 점자 인식. 국제공학학회지, 19192, 2018. 2

[18] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. 한 번만 보세요: 통합된 실시간 물체 감지. 컴퓨터 비전 및 패턴 인식에 관한 IEEE 회의 진행, 페이지 779-788, 2016. 3, 4 [19] Joseph Redmon 및 Ali Farhadi. Yolo9000: 더 좋고, 더 빠르고, 더 강력합니다. 컴퓨터 비전 및 패턴 인식에 관한 IEEE 회의 진행, 페이지 7263-7271, 2017. 3, 4 [20] Joko Subur, Tri Arief Sardjono 및 Ronny Mardiyanto.

윤곽 찾기와 인공 신경망을 이용한 점자 문자 인식. JAVA Journal of Electrical and Electronics Engineering, 14(1), 2016. 3 [21] V Udayashankara et al. 엠보싱 점자 문자의 광학 인식을 위한 소프트웨어 알고리즘에 대한 검토. 국제 컴퓨터 응용 저널, 81(3):25-35, 2013. 2

[22] 가야트리 베누고팔-와이라가데. 카메라 지원 스마트폰을 이용한 점자 인식. Int J Eng Manuf, 4:32-39, 2016. 2, 3 [23] Paul Viola 및 Michael Jones. 간단한 기능의 강화된 캐스케이드를 사용하여 신속한 물체 감지. 컴퓨터 비전 및 패턴 인식에 관한 2001 IEEE 컴퓨터 학회 회의 절차에서. CVPR 2001, 볼륨 1, 페이지 I-I.

IEEE, 2001. 3

[24] Shanjun Zhang 및 Kazuyoshi Yoshino. 카메라가 내장된 휴대폰의 점자 인식 시스템. Innovative Computing, Informatio and Control(ICICIC 2007)에 관한 두 번째 국제 회의, 223-223페이지. IEEE, 2007. 2, 3