ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2020 June: 26(2): 1~25 http://dx.doi.org/10.13088/jiis.2020.26.2.001

관심 문자열 인식 기술을 이용한 가스계량기 자동 검침 시스템*

이교혁
연세대학교 기술경영학협동과정
(kyohyuk.lee@yonsei.ac.kr)

김태연
한국과학기술원 컴퓨터과학과
(kty5177@kaist.ac.kr)

김우주
연세대학교 정보산업공학과
(wkim@yonsei.ac.kr)

본 연구에서는 모바일 기기를 이용하여 획득한 가스계량기 사진을 서버로 전송하고, 이를 분석하여 가스 사용량 및 계량기 기물 번호를 인식함으로써 가스 사용량에 대한 과금을 자동으로 처리할 수 있는 응용 시스템 구조를 제안하고자 한다. 모바일 기기는 일반인들이 사용하는 스마트 폰에 준하는 기기를 사용하였으며, 획득한 이미지는 가스 공급사의 사설 LTE 망을 통해 서버로 전송된다. 서버에서는 전송받은 이미지를 분석하여 가스계량기 기물 번호 및 가스 사용량 정보를 추출하고, 사설 LTE 망을 통해 분석 결과를 모바일 기기로 회신한다. 일반적으로 이미지 내에는 많은 종류의 문자 정보가 포함되어 있으나, 본 연구의 응용분야인 가스계량기 자동 검침과 같이 많은 종류의 문자 정보 중 특정 형태의 문자 정보만이 유용한 분야가 존재한다. 본 연구의 응용분야 적용을 위해서는 가스계량기 사진 내의 많은 문자 정보 중에서 관심 대상인 기물 번호 및 가스 사용량 정보만을 선별적으로 검출하고 인식하는 관심 문자열 인식 기술이 필요하다. 관심 문자열 인식을 위해 CNN (Convolutional Neural Network) 심층 신경망 기반의 객체 검출 기술을 적용하여 이미지 내에서 가스 사용량 및 계량기 기물번호의 영역 정보를 추출하고, 추출된 문자열 영역 각각에 CRNN (Convolutional Recurrent Neural Network) 심층 신경망 기술을 적용하여 문자열 전체를 한 번에 인식하였다. 본 연구에서 제안하는 관심 문자열 기술 구조는 총 3개의 심층 신경망으로 구성되어 있다. 첫 번째는 관심 문자열 영역을 검출하는 합성곱 신경망이고, 두 번째는 관심 문자열 영역 내의 문자열 인식을 위해 영역 내의 이미지를 세로 열 별로 특징 추출하는 합성곱 신경망이며, 마지막 세 번째는 세로 열 별로 추출된 특징 벡터 나열을 문자열로 변환하는 시계열 분석 신경망이다. 관심 문자열은 12자리 기물번호 및 4 ~ 5 자리 사용량이며, 인식 정확도는 각각 0.960, 0.864 이다. 전체 시스템은 Amazon Web Service 에서 제공하는 클라우드 환경에서 구현하였으며 인텔 제온 E5-2686 v4 CPU 및 Nvidia TESLA V100 GPU를 사용하였다. 1일 70만 건의 검침 요청을 고속 병렬 처리하기 위해 마스터-슬레이브 처리 구조를 채용하였다. 마스터 프로세스는 CPU 에서 구동되며, 모바일 기기로 부터의 검침 요청을 입력 큐에 저장한다. 슬레이브 프로세스는 문자열 인식을 수행하는 심층 신경망으로써, GPU 에서 구동된다. 슬레이브 프로세스는 입력 큐에 저장된 이미지를 기물번호 문자열, 기물번호 위치, 사용량 문자열, 사용량 위치 등으로 변환하여 출력 큐에 저장한다. 마스터 프로세스는 출력 큐에 저장된 검침 정보를 모바일 기기로 전달한다.

주제어 : 가스계량기, 자동 검침, 선택적 문자 인식, 합성곱 신경망, 순환 신경망, 고속 병렬처리

논문접수일 : 2020년 4월 11일 논문수정일 : 2020년 5월 11일 게재확정일 : 2020년 5월 14일
원고유형 : 일반논문 교신저자 : 김우주

* 본 논문은 과학기술정보통신부 및 정보통신산업진흥원의 '고성능 컴퓨팅 지원' 사업으로부터 지원받아 수행하였음.

1. 개요

본 연구는 가스계량기 사진을 분석하여 가스 사용량 및 계량기 고유번호를 인식함으로써 가스 사용량에 대한 과금을 자동으로 처리할 수 있는 응용 시스템 구축에 관한 것이다. 최근 LoRa, NB-IoT, LTE-M 등과 같은 저전력 광대역 무선통신 표준 기술이 개발되어 다양한 분야의 machine-to-machine 응용 분야에 적용되고 있다. 가스계량기에 저전력 광대역 무선통신 송신 모듈을 탑재하고 서버에 수신 기능을 구현하면 월별 사용량에 대한 자동 검침을 가능하게 할 수 있으나, 이를 위해서는 모든 개별 가스 소비자가 저전력 광대역 무선통신 모듈이 탑재된 가스계량기를 구매하여야 한다 (국내 가스계량기는 소비자 소유이다). 가스 공급사가 저전력 광대역 무선통신 모듈이 탑재된 가스계량기를 모든 소비자에게 공급하는 것 또한 막대한 재정이 필요하므로 현실적으로 어려움이 있다. 또한 저전력 광대역 무선 통신 모듈은 배터리로 구동 되므로 주기적으로 배터리를 교체해주지 않을 시 검침이 이루어지 않으므로 과금에 문제가 발생하게 된다. 이러한 현실적인 이유로 대부분의 가스 공급사는 검침원 운영을 통해 매월 직접 가스계량기가 설치된 장소를 방문하여 나안으로 사용량을 확인한 후, 모바일 기기 등에 수작업으로 사용량을 입력하는 방식으로 검침 및 과금을 수행하고 있으며, 이와 같은 수작업 방식에는 매우 많은 시간이 소요된다.

가스계량기를 촬영한 이미지를 분석하여 계량기 고유번호 및 사용량 문자를 자동 인식하고 가스 사용량을 자동 입력할 수 있는 시스템을 구축할 수 있다면, 가스 공급사의 검침 업무 효율 향상을 기대할 수 있다. 이를 위해서는 가스계량기 이미지를 분석하여 기기의 고유번호 및 사용량을 인식하는 문자 인식 기술 및 여러 검침원이 동시에 전송하는 이미지를 빠르게 분석하는 대용량 병렬 컴퓨팅 기술이 필요하다.

문자 인식 기술은 이미지 내의 문자 영역을 검출하고, 그것이 어떤 문자인지를 판별하는 기술로써, 초기에는 전자 파일을 인쇄한 이미지 내의 정형화된 문자들을 인식하는 것에 집중되었으나, 최근에는 일반 사진 이미지에 나타나는 비정형 문자 인식에 기술이 집중되고 있다 (Jaderberg, 2016). 전자 파일 인쇄에 나타나는 문자는 흑백 이미지로 구성되며 특정 폰트의 문자가 가로로 나열된 것이 대부분이나, 일반 사진 이미지 속의 문자는 다양한 컬러와 폰트로 구성되어 있을 뿐 아니라 문자의 위치 또한 다양한 방향으로 나타난다. 문자 인식 기술은 이미지에 내포된 의미를 파악하여 환경에 대한 이해를 가능하게 하여 인터넷 상의 무수히 많은 이미지에 대한 태그 정보 자동 생성 및 검색, 비디오 영상에 대한 텍스트 설명 자동 생성, 자율 주행차가 운행하는 주변 환경 분석 등 다양한 분야에 응용될 수 있다 (Liao, Shi, & Bai, 2018; Liu, Chen, Wong, Su, & Han, 2016; Tian, Lu, & Li, 2017; Zhou et al., 2017).

문자 인식 기술은 이미지에 포함된 모든 문자를 인식하는 것이 일반적이나, 이미지에 포함된 다양한 문자들 중에서 관심 있는 특정 문자만을 인식해야하는 응용 분야도 존재한다. 예를 들면, 가스, 전기 및 수도 등 사용량을 표시하는 검침 계량기의 과금 대상 숫자 인식, 제품의 고유 번호 인식, 신용카드의 카드번호 및 유효 기간 인식 등이 그것이다. 일반적인 문자 인식 기술은 이미지 내의 모든 문자 영역을 검출하고 인식하므로, 이러한 관심 대상 문자만을 검출하고 인식

하는 분야에는 적합하지 않다. 관심 대상 문자 인식을 위해서는 기울어짐, 측면, 사선, 원거리, 곡선 등 다양한 문자 배치에 대해 안정적으로 인식하여야 할 뿐 아니라, 문자로 추정되는 영역이 관심 대상 문자인가를 판별하는 것이 필요하다.

용어의 통일을 위해 독립적으로 존재하는 하나의 개별 문자를 단위 문자라고 정의하고, 단위 문자들이 연속적으로 나열되어 하나의 의미를 나타내는 것을 문자열이라고 정의한다.

2. 관련 연구

2.1 심층 신경망

이미지 분석 기술은 영상분석 전문가에 의해 정의된 특징 값 및 통계 분석 기술에 기반을 둔 분석 방법과 데이터에 기반을 두어 신경망 가중치를 학습시키는 기계학습 방법으로 분류할 수 있다.

전문가 지식에 의한 이미지 분석은 이미지를 잘 표현할 수 있는 특징 값 변환 방식을 미리 정의하고, 입력된 이미지를 특징 값으로 변환한 후 특징 값의 통계적 특성을 이용하여 패턴을 분석하는 방법이다. 이러한 방식에는 각 픽셀 위치에서의 픽셀 강도의 기울기 분포를 참조하여 패턴을 분석하는 histogram of gradient 방법 (Dalal & Triggs, 2005), 여러 단계 크기의 이미지를 Gaussian 필터링 처리 후 차분 영상을 만들고 각 픽셀에서의 edge 강도 및 상대적인 기울기 분포 정보를 바탕으로 패턴을 분석하는 scale invariant feature transform (Lindeberg, 2012; Lowe, 2004), 원점을 기준으로 각 픽셀을 지나는 edge 선들을 거리, 각도 등의 정보로 변환한 특징 값을 이용하여 패턴을 분석하는 hough transform (Ballard, 1981) 등이 있다.

반면, 기계학습 방식의 영상 분석은 이미지를 분석하는 연산의 구조만을 먼저 설계한 후, 연산의 가중치들은 데이터로부터 학습하는 방식이다. 연산의 구조는 인간 뇌를 구성하는 단위 신경세포를 모방한 perceptron을 기본 단위로 한다. Perceptron은 여러 개의 입력 신호에 가중치를 곱하여 합산한 후, 임계치와 비교하여 출력 여부를 결정하는 비선형 활성화 함수를 통해 하나의 출력을 내보낸다. Perceptron의 가중치와 임계치 파라메터는 학습 데이터를 바탕으로 오류를 최소화하는 방향으로 체험적 업데이트를 통해 학습시킨다 (Rosenblatt, 1958). 신경망 기술은 여러 perceptron의 입력, 출력을 상호 연결하여 인간의 뇌 구조를 모방한 연산 구조를 가지고 있다. Perceptron 개수가 적으면 데이터 기반의 체험적 업데이트를 통해 파라메터를 학습할 수 있으나, perceptron 의 개수가 많아지면 체험적 업데이트로는 정확한 파라메터를 추정하는 것이 어려워지는 문제가 있다. 이러한 문제를 해결하기 위해, 경사 차분 기반의 오류 하방 전파를 통해 학습 데이터의 오류값을 정확하게 추정하는 방식이 개발되었다 (Rumelhart, Hinton, & Williams, 1986). 경사 하방 전파 방식은 최종 출력단의 오류값에 대한 신경망 내 모든 파라메터의 편미분을 수학적으로 계산하는 기술로써, 이를 통해 파라메터 값의 변화가 오류값에 미치는 영향을 정확하게 추정함으로써 학습 데이터에 최적화된 신경망 파라메터를 찾을 수 있게 해준다.

경사 하방 전파 기술에 의해 파라메터를 정확하게 추정할 수 있게 되었으나, 학습 과정에서 파라메터 값이 수렴하고 신경망 층수가 늘어나면서 경사 값이 0으로 수렴하면서 학습이 이루

어지지 않거나, 발산하게 되는 문제가 발생할 수 있다. 이러한 문제를 해결하기 위해 신경 노드 출력값 정규화 (Ioffe, 2015), ReLU 비선형 활성 함수 (Nair & Hinton, 2010) 및 파라메터 초기값 정규화 (Glorot & Bengio, 2010) 등의 기술들이 제안되었으며, 이로 인해 경사 값 수렴 문제를 완화시키고 신경망 층수를 기하급수적으로 증가시키면서 다양한 구조의 심층 신경망 기술이 제안되었다.

심층 신경망의 신경망 층 개수의 증가와 함께 연산 복잡도도 그에 비례하여 증가하였다. 현재까지 제안된 심층 신경망의 파라메터 수는 수천만 개 ~ 수억 개에 달하며 (Canziani, 2016), 이러한 심층 신경망의 파라메터를 안정적으로 학습시키기 위해서는 많은 양의 데이터가 필요하다. 인터넷, 소셜 네트워크 서비스 및 하드웨어 기술의 발달과 함께 심층 신경망 학습을 위한 이미지 데이터 확보가 종전보다 용이해 졌으나, 수많은 심층 신경망 파라메터를 안정적으로 학습시키는 것에는 한계가 있다. 심층 신경망 파라메터를 학습시키기 위해서는 파라메터 수에 비례하는 많은 양의 학습 데이터가 필요하다. 제한된 양의 학습 데이터를 이용하여 심층 신경망 파라메터를 학습시키기 위해서 이미지 회전, 이미지 반전, 음영 조정, 색감 조정, 주성분 분석 등의 다양한 방법들을 통해 주어진 학습 데이터 양을 인위적으로 증가시키는 기술들이 제안되었다 (Krizhevsky, 2012).

2.2 이미지 분류

공간 기하학적 유사성을 해석하는 인간의 시각 인지 기능을 모방한 신경망 구조인 neocognitron (Fukushima, 1980)이 개발된 이 후, 이를 발전시켜 작은 영역을 분석한 정보를 결합하여 보다 큰 영역을 해석해 나가는 인간 시각 지능의 구조적 해석 과정을 모방한 합성곱 신경망 구조가 제안되었다 (LeCun, Bottou, Bengio, & Haffner, 1998). 합성곱 신경망 구조는 국지 영역을 분석하는 합성곱 커널 분석층을 반복 계층적으로 쌓고, 분석층의 해상도를 줄여나가는 과정을 통해 합성곱 커널의 영상 해석 영역을 점진적으로 넓혀 나가는 다층 신경망 구조이다. 합성곱 신경망은 분석층의 구조를 어떻게 구성하느냐에 따라 서로 다른 특성 및 성능을 얻을 수 있다. LeCun et al. (1998)이 제안한 초기의 합성곱 신경망은 22개 층으로 구성되었으나, 그 후 다양한 구조의 합성곱 신경망이 제안되면서 신경망 층의 개수를 기하급수 적으로 증가시키면서 심층 신경망으로 발전하였다. AlexNet (Krizhevsky, 2012), ZFNet (Zeiler, 2014), GoogLeNet (C. Szegedy, Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A., 2015), Inception (C. Szegedy, Ioffe, S., Vanhoucke, V., & Alemi, A. A., 2017), ResNet (K. He, Zhang, X., Ren, S., & Sun, J., 2016), VGG (Simonyan, 2014), DenseNet (G. Huang, Liu, Van Der Maaten, & Weinberger, 2017), FractalNet (Larsson, 2016), Network-in-Network (Lin, 2013) 등은 초기 합성곱 신경망 대비 분석층의 개수가 기하급수적으로 늘어난 새로운 심층 신경망 구조로써, 이들 심층 합성곱 신경망들은 합성곱 커널의 기본 구조, 전체 신경망 층 개수, 각 신경망 층의 해상도, 신경망 층간의 연결 구조 등에 따라 영상 분석 특성을 가지며, 구조에 따라 연산 복잡도도 달라진다.

AlexNet (Krizhevsky, 2012) 은 입력 영상을 처리하는 합성곱 신경망을 2개의 병렬 구조로 분리한 뒤, 최종 특징 벡터 추출 단계에서 다시 하

나의 합성곱 레이어로 합치는 구조를 적용하였다. AlexNet 개발 당시의 GPU 하드웨어 성능 미흡으로 인해 AlexNet은 두 개의 분리된 합성곱 신경망 구조를 채택하였으나, 그 이후 GPU 하드웨어 성능 개선과 함께 다양한 단일 합성곱 신경망 계층 구조가 개발되었다. VGG (Simonyan, 2014) 구조는 3x3 합성곱 커널의 신경망 층 및 2x2 픽셀 영역 최대값 출력층을 반복 계층적으로 배치하는 간단한 합성곱 신경망 구조만으로도 이미지 분류의 성능을 크게 향상시킬 수 있음을 보였다. GoogLeNet (C. Szegedy, Ioffe, S., Vanhoucke, V., & Alemi, A. A., 2017)은 패턴 분석을 위한 3x3, 5x5 합성곱 커널 및 연산량 감축을 위한 1x1 병목 합성곱 커널로 구성되는 모듈을 기본 분석 단위로 하며, 기본 분석 모듈을 순차 계층적으로 연결함으로써, 다양한 크기의 국지 영역 분석이 가능하면서 전체적인 연산량을 저감시킬 수 있는 구조를 제시하였다. 또한 GoogLeNet은 심층 신경망의 최종 출력층 뿐만이 아니라 중간 신경망 층에도 손실 함수를 삽입하는 방법을 적용하여, 경사 하방 전파 시 문제가 될 수 있는 경사 차분 사라짐 문제를 완화시켰다. ResNet (K. He, Zhang, X., Ren, S., & Sun, J., 2016)은 하위 신경망 층의 출력을 차상위 신경망 층의 출력과 합치는 구조로써, 하위 신경망 층의 정보를 차상위 계층으로 우회시키는 특성을 이용하여 하위 신경망 층에서 이미 학습된 정보를 차상위 신경망 층에서 부가적으로 학습하는 것을 방지하고, 경사 하방 전파 경로를 단축함으로써 경사 차분 사라짐 문제도 완화시킬 수 있는 구조이다. 또한 이러한 구조는 차상위 신경망 층이 학습해야 할 정보량을 줄여서 보다 학습이 안정적으로 이루어지게 되는 효과가 있다. DenseNet (F. Iandola, Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., & Keutzer, K., 2014)은 하위 신경망 층 출력을 모든 상위 신경망 층 출력과 통합하는 구조로 설계되었으며, 이러한 구조는 하위 신경망 층에서 분석된 높은 해상도의 분석 특징들을 상위 층으로 전달함으로써 다양한 크기의 패턴 분석을 가능하게 한다. FractalNet (Larsson, 2016)은 하나의 합성곱 분석층을 복수개의 합성곱 분석층과 병치시킨 기본 처리 단위를 계층적으로 쌓아 올린 구조로써, 하나의 기본 구성단위 내에 여러 종류의 네트워크 연결 패턴이 포함되어 있으므로, 신경망 계층의 입력에서 출력으로 이어지는 경로를 다양하게 선택할 수 있다. 이는 ResNet과 같은 차분 경로 구조 특성을 유지하면서, GoogLeNet과 같은 중간 레이어 손실 삽입 없이도, 짧은 경사 하방 전파 경로 선택이 가능하므로 경사 차분의 하방 전파를 용이하게 하여 안정적인 심층 신경망 학습을 가능하게 한다.

AlexNet 이 후, 다양한 심층 신경망 구조 기술들이 제안됨에 따라 이미지 내의 객체 종류를 분류하는 한정된 영상 분석 분야에서 종전의 전문가 지식 및 통계 기술 대비 월등한 성능을 얻을 수 있게 되었다 (Canziani, 2016).

2.3 객체 검출

이미지 분류 심층 신경망 기술들은 이미지 내의 객체 종류 뿐 아니라, 객체의 위치 및 영역 정보도 함께 분석하는 객체 검출 기술로 확대 적용되었다. 이미지 내에 객체의 위치, 영역 정보를 추정하기 위해서는 각 픽셀이 객체 영역에 포함되는지를 판별하여야 한다. Gower (1969)는 각 픽셀 간의 연결을 하나의 트리 구조로 해석하는 minimum spanning tree 방식을 이용하여 이미지 내의 객체 영역을 분할하는 기술을 제시하였다.

Spanning tree는 인접하는 픽셀 간의 연결성 지수를 이용하여, 모든 픽셀들 간의 연결 정도를 tree 구조로 나타낸 것으로써, minimum spanning tree는 spanning tree 중 픽셀 간 거리의 총 합이 가장 작은 것을 동일 객체를 구성하는 픽셀 집합으로 정의하는 것이다. 이를 발전시켜 Felzenszwalb (2004)는 픽셀 간의 연결성 지수와 함께 동일 객체로 추정되는 영역 내부 거리 및 서로 다른 객체 영역 간의 상대적 거리를 정의하고 두 종류의 거리를 비교하여 서로 다른 객체 영역의 경계선을 결정하는 그래프 기반 이미지 분할 (graph-based image segmentation) 기술을 제안하였다. Uijlings (2013)은 minimum spanning tree 및 그래프 기반 이미지 분할 기술을 통합하고, 이에 더하여 컬러 유사도, 질감 유사도, 크기 유사도 및 영역 유사도의 네 가지 유사도를 하나의 유사도 지표로 통합하여 객체를 인식하는 선택적 검색 방식을 제시하였다.

심층 신경망 기술의 성능이 향상되면서, 객체 검출 분야에도 심층 신경망 기술이 적용되기 시작했는데, Girshick (2014)은 그래프 기반 이미지 분할 (Felzenszwalb, 2004) 및 선택적 검색 방식 (Uijlings, 2013)을 이용하여 2,000개의 객체 영역 후보를 생성하고, 이들 후보 영역을 대상으로 심층 합성곱 신경망을 적용하여 각 후보 영역에 대한 객체 존재 확률을 추정한 후, 중첩 후보 영역을 제거하는 방식의 R-CNN (Regions with Convolutional Neural Network features) 기술을 제안하였다. Girshick (2015)은 R-CNN 의 문제점인 객체 후보 영역의 크기를 심층 합성곱 신경망의 입력 해상도로 변환하는 과정에서 발생하는 이미지 왜곡을 보완하기 위해 최종 특징 맵을 등분할 하여 영역 최대값 출력층을 구성하는 Spatial Pyramid Pooling 방식 (K. He, Zhang, Ren, & Sun, 2015)을 적용하였다. 이 후, Ren (2015)은 객체 후보 영역을 생성하는 별도의 심층 신경망을 추가하여, 객체를 인식하는 심층 신경망과 함께 2개의 심층 신경망으로 구성된 Faster R-CNN 기술을 제안하였다. 이러한 R-CNN 계열의 기술들은 객체 후보 영역을 추정하는 부분과 객체를 인식하는 부분으로 나뉘어져 있어 학습 시 2 개 모듈을 순차적으로 학습해야 하며, 추론 시 2 개의 심층 신경망 연산을 수행해야 하므로 처리 속도가 늦은 단점이 있다.

이러한 R-CNN 계열 기술들의 단점을 극복하기 위해 단일 심층 신경망을 이용하여 객체 영역 후보 추정과 객체 인식을 동시에 처리하는 기술이 제안되었다. Liu, Anguelov, et al. (2016)는 입력 영상을 등분할 하고 각 등분할 영역 위치에서 가로 세로 비율이 서로 다른 기저 사각형 (anchor box) 영역을 생성한 후 기저 사각형 영역들을 실제 객체가 위치하고 있는 영역들로 할당하는 SSD (Single Shot multi-box Detection) 기술을 제안하였다. SSD 기술은 기저 사각형의 위치 및 크기와 검출하고자 하는 객체의 위치 및 크기의 상대적 차이 정보를 추정한다. 또한, 가로 세로 비율이 다른 기저 사각형 별로 별도의 합성곱 커널 필터를 할당하였으며, 다양한 크기의 객체 검출 성능 향상을 위해 하위 신경망 층의 특징 맵 정보를 객체 추정 시 활용하는 구조를 가진다. SSD는 기존에 존재하는 VGG (Simonyan, 2014) 심층 신경망 구조에 기반하여 객체 검출에 필요한 추가 정보를 추정할 수 있도록 수정하여 설계를 하였다. 반면, J. Redmon, Divvala, Girshick, and Farhadi (2016)는 GoogLeNet (C. Szegedy, Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A., 2015)을 변형하여 객체 검출에 특화된 독자적인 YOLO (You Look Only

Once) 심층 신경망 구조를 제안하였다. YOLO는 24개의 합성곱 신경망 층과 2 개의 fully-connected layer를 가진다. YOLO 기술은 이미지를 등분할 하고, 등분할 된 각 기저 사각형 당 복수개의 객체 추정 정보를 출력하는 구조를 가지고 있다. 또한 객체 위치 정보와 기저 사각형과의 상대적인 거리를 0 ~ 1 값으로 표준화하여 객체 추정의 안정성을 향상시켰다. 이 후, J. Redmon, & Farhadi, A. (2017)는 YOLO 를 발전시킨 YOLOv2 구조를 제안하였다. YOLOv2 구조는 이미지 등분할 개수를 홀수 개로 설정하여 이미지 중앙에 항상 기저 사각형 1개가 배치될 수 있도록 하고, 기저 사각형의 개수를 약 20배 이상 증가시켰으며, 학습 데이터의 객체 영역 정보를 K-Means 방법으로 군집화하여 학습 데이터에 최적화된 기저 사각형의 가로 세로 비율을 설정하고, 기저 사각형 1개 당 5개의 객체 후보 영역을 출력하는 구조를 가지고 있다. 또한 SSD와 유사하게 하위 신경망 층의 특징 맵 정보를 상위 신경망 층의 특징 맵 정보와 융합함으로써 다양한 크기의 객체를 정확하게 추정할 수 있다. YOLOv2는 1x1 합성곱 신경망 층을 적절히 배치하여 전체적인 연산량을 저감시켰으며, 서로 다른 학습 데이터의 레이블 구조를 하나의 레이블 구조로 통합한 WordTree 형식의 계층적 레이블 구조를 제시하였다. 이 후 J. Redmon, & Farhadi, A. (2018)는 YOLOv2를 추가적으로 개선하여 YOLOv3 구조를 제시하였다. YOLOv3는 객체 할당이 되지 않은 기저 사각형이 출력하는 위치 정보를 손실 함수에 반영하지 않고, 객체 존재 여부에 대한 확률 값만을 손실 함수에 반영한다. 객체 존재 확률에 대한 최종 출력 신경망을 Softmax 함수를 사용하지 않고 독립적인 Logistic 함수를 사용함으로써 하나의 기저 사각형이 여러 개의 객체와 매핑될 수 있도록 하였다. 이렇게 함으로써 복수개의 객체가 겹쳐서 위치하고 있는 경우에 대한 검출력을 향상시킬 수 있다.

2.4 문자 인식

문자 인식 기술은 이미지 내 어느 위치에 문자가 존재하는가를 검색하는 문자 영역 검출 기술 (Liao et al., 2018; Liao, Shi, Bai, Wang, & Liu, 2017; B. Shi, et al., 2017; Tian et al., 2017; Zhou et al., 2017)과 검출된 영역 내에 포함된 문자들의 종류가 무엇인지를 판별하는 문자열 인식 기술 (Jaderberg, 2016; B. Shi, Bai, & Yao, 2017; Zhu, 2017)로 구성된다. 이미지 내에는 여러 위치에 다양한 형태의 문자가 위치할 수 있으며, 일반적으로 문자 인식 기술은 이러한 모든 문자를 검출하고 인식한다.

경사 하방 전파 기술에 의한 신경망 기술 발전은 문자 인식 분야에 처음으로 성공적으로 적용되었다. 신경망을 이용한 문자 인식은 이미지 내에 존재하는 모든 단위 문자 영역을 검출하고, 검출된 국지 이미지 영역에 신경망 기술을 적용한 분류기를 적용하여 단위 문자를 인식하는 구조를 가진다 (Fukushima, 1980; LeCun et al., 1998). 단위 문자 인식 학습을 위해서는 이미지 내의 모든 단위 문자에 대한 위치 및 문자 종류 레이블을 부여하여야 한다. 단위 문자별로 레이블을 부여하는 것은 학습을 위한 레이블 데이터 생성에 많은 시간과 노력이 필요하다. 이러한 단점을 극복하기 위하여 Tian et al. (2017)는 준 지도 학습 (semi-supervised learning) 및 약한 지도 학습 (weakly supervised learning) 방식을 제안하였다. 준 지도 학습은 작은 크기의 데이터를 이용하여 지도 학습을 적용한 가벼운 지도 학습 모

델을 생성하고, 지도 학습에 의해 생성한 모델을 이용하여 레이블이 없는 대용량 데이터를 인식한 후, 이들 중 신뢰도가 높은 추정치들을 새로운 단위 문자 레이블로 간주하여 지도 학습을 위한 대용량 데이터 레이블을 생성하는 방식이다. 이렇게 함으로써, 단위 문자 인식에 필요한 레이블 데이터 생성 시간 및 노력을 줄이면서도, 데이터 부족에 의한 성능 하락을 일정 부분 보완할 수 있다. 약한 지도 학습은 단위 문자별 레이블을 생성하지 않고, 문자열 단위의 레이블을 생성한 뒤, 가벼운 지도 학습 모델을 이용하여 문자열 단위 레이블 내의 단위 문자를 인식하고, 이들 중 신뢰도가 높은 추정치들을 새로운 단위 문자 레이블로 간주하여 지도 학습을 위한 대용량 데이터 레이블을 생성하는 방식이다. 약한 지도 학습은 지도 학습과 준 지도 학습의 중간 형태의 학습 방법으로써, 준 지도 학습 보다 더 좋은 성능을 얻을 수 있으나, 약한 레이블 생성을 위한 시간과 노력이 일정 부분 요구되는 단점이 있다.

단위 문자에 대한 검출 및 인식이 완료된 후, 문자열을 인식하기 위해서는 단위 문자들을 조합하여 의미 있는 하나의 문자열 그룹으로 합치는 것이 필요하다. 이를 위해 Baek, Lee, Han, Yun, and Lee (2019)은 인접한 2 개의 단위 문자 영역의 중심을 포함하는 추가적인 사각형 영역을 추정하여 단위 문자들의 연결성을 모델링하고, 연결성이 높은 단위 문자들을 조합하여 문자열을 구성하는 기술을 제안하였다. 즉, 2 개의 인접한 단위 문자 경계 영역을 별도의 영역으로 지정하고, 영역 정보를 학습함으로써 단위 문자의 연결 정도를 추정하여 별도의 문자열 군집화를 하지 않고도 문자열을 구성할 수 있다.

B. Shi et al. (2017)은 합성곱 신경망과 순환 신경망을 조합하여 문자열 영역을 한번에 해석하고 인식할 수 있는 CRNN (Convolutional Recurrent Neural Network) 구조 기술을 제안하였다. CRNN은 먼저 합성곱 신경망을 이용하여 이미지의 공간적 패턴을 분석한 합성곱 특징 맵을 생성하고, 합성곱 특징 맵을 열 단위 벡터로 분할하고, 분할된 특징 벡터를 좌측에서 우측 순서로 순환 신경망에 순차적으로 입력한다. 순환 신경망은 공간적 해석 정보를 담고 있는 열 단위의 각 특징 벡터를 단위 문자로 변환하여 출력한다. 순환 신경망은 시계열 데이터 분석을 위한 신경망으로써, 숨은 신경망 층 노드 (hidden layer node)의 자기 순환 루프를 통해 기준 시점과 다음 시점의 네트워크를 연결시킴으로써, 이전 시점의 정보를 시간의 흐름에 따라 다음 시점으로 전달하는 구조로 이루어져 있다. 그러나, 시점이 늘어날수록 신경망 층이 깊어지면서 경사 차분 사라짐 문제가 발생한다. 이를 보완하기 위해 Hochreiter (1997)는 숨은 신경망 노드에 forget gate 를 삽입한 Long Short-Term Memory (LSTM) 구조를 제안하였다. CRNN은 이전 시점과 다음 시점의 숨은 신경망 노드가 상호 연결되어 서로 다른 시점의 정보를 상호 교환할 수 있는 Bi-LSTM (Z. Huang, Xu, & Yu, 2015) 구조를 채용하였다.

3. 연구 방법

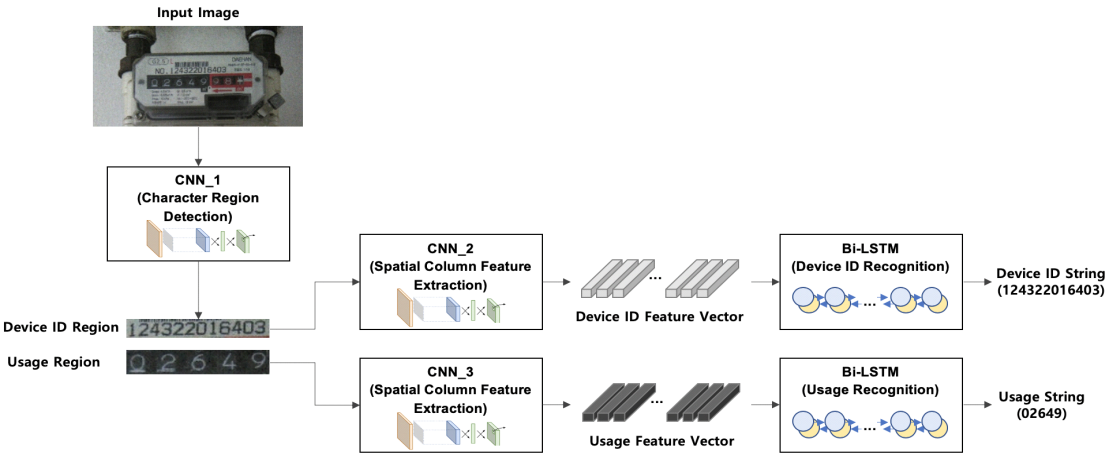
본 연구에서는 이미지 내에 존재하는 다양한 문자열들 중에서 관심 대상 문자열만을 선별하여 인식할 수 있는 관심 문자열 인식 기술 구조를 제시하고자 한다. 가스계량기 자동 검침을 위해서는 가스계량기 이미지 내의 기기 고유번호 문자열 및 사용량 문자열을 인식해야 한다. 계량기 이미지 내에는 관심 대상 문자열인 고유번호,

사용량 외에 제조사명, 모델명 및 각종 스펙 관련 문자열들이 다수 포함되어 있으며, 형태, 위치 및 내용이 제조사별로 각기 다르다. 기기 고유번호 문자열은 12자리 숫자로 구성되어 있으며, 사용량 문자열은 4 ~ 5자리 숫자로 구성되어 있다.

단위 문자 검출 기반 방식을 이용하여 관심 대상 문자열을 구별하기 위해서는 단위 문자를 검출하고 조합하여 문자열을 생성한 후, 검출된 단위 문자 또는 문자열에 대해 관심 문자열 대상 여부를 판별해야 한다. 일반적으로 문자열은 복수개의 단위 문자로 구성되며, 단위 문자 수가 많을수록 관심 대상 문자열 검출 시 오류 전파의 문제가 발생할 수 있다. 즉, 복수개의 단위 문자 중 하나의 단위 문자라도 검출 오류가 발생하면 최종 문자열 검출에 오류가 발생하게 된다. 또한, 관심 문자열 영역을 검출한 이 후, 단위 문자를 그룹화하여 관심 대상 문자열을 인식해야 하는 등 여러 단계의 검출 과정이 순차적으로 연결되어야 하므로 전체적인 성능 저하의 원인이 될 수 있다 (Zhou et al., 2017). 더불어 단위 문자 기반의 검출 방식은 학습용 데이터를 모든 단위 문자별로 레이블을 생성해야 하므로 학습용 데이터를 준비하는데 시간과 비용이 많이 드는 단점이 있다.

본 연구에서는 전체 문자열 영역을 직접 추정하고, 전체 관심 문자열을 한 번에 인식할 수 있는 방식을 채택하고자 한다. 이를 위해서, 이미지 내에 존재하는 문자열 중 관심 대상인 문자열인 기기 고유번호 및 사용량에 대해서만 위치 정보 및 문자열 정보를 데이터 레이블에 포함시켰다. 관심 문자열을 둘러싸는 가장 작은 직사각형 영역의 4개 꼭짓점 좌표를 위치 정보 레이블로 부여하고, 직사각형 영역 내의 문자 정보들을 문자열 레이블로 부여하였다.

본 연구의 전체적인 기술 구조를 <Figure 1>에 나타내었다. 먼저, 객체 검출 기술을 적용하여 관심 문자열이 위치하는 영역 전체를 한 번에 검출하고, 검출된 관심 문자열 영역 내에 존재하는 복수개의 단위 문자를 인식하는 2 단계 심층 신경망 구조로 구성하였다.



〈Figure 1〉 Selective Optical Recognition Process

3.1 관심 문자열 영역 검출

관심 문자열 영역을 검출하기 위한 방법으로는 심층 신경망 기반의 객체 검출 기술을 적용할 수 있다. 최근의 객체 검출 기술은 크게 2 가지로 분류할 수 있다. 첫 번째 방식은 복수개의 객체 위치 후보를 검색한 후, 각 후보 위치에서 검출하고자 하는 객체의 존재 여부를 판별하는 2 단계 처리 구조의 기술들이다 (Girshick, 2015; Ren, 2015; Uijlings, 2013). 2 단계 처리 구조의 기술들은 minimum spanning tree (Gower, 1969), graph-based segmentation (Felzenszwalb, 2004) 등을 이용하여 객체 영역 후보들을 생성하고, 생성된 각 후보 영역들에 대해 객체 존재 확률을 추정하고 객체의 종류를 분류하는 구조로써, 객체 영역을 보다 정확하게 추정할 수 있는 장점이 있으나, 연산복잡도가 큰 단점이 함께 존재한다. 두 번째 방식은 이미지 영역을 미리 정해진 크기의 영역들로 등분할 하고, 등분할 영역 (anchor box) 각각을 검출하고자 하는 객체에 할당하여 anchor box 와 객체 영역의 정합도를 측정하고 객체의 종류를 분류하는 1 단계 처리 방식이다 (Liu, Anguelov, et al., 2016; J. Redmon, & Farhadi, A., 2018; J. Redmon et al., 2016). 1 단계 처리 방식의 기술들은 객체 영역에 대한 사전 검출 과정이 없으므로 연산 복잡도가 낮고, 학습 과정이 단순한 장점이 있다.

본 연구의 응용분야는 다수의 검침원이 전체 가스계량기를 주어진 시간 내에 처리해야 하는 응용분야로써, 1일 처리 요청 건수가 약 70만 건에 이르며, 이미지 문자 인식 요청에 대한 응답 속도는 1초 이내여야 한다. 이는 1초당 평균적으로 약 24건의 이미지를 처리해야 하며, 최대 응답 속도가 1초를 넘지 않아야 한다. 문자 인식 기술의 복잡도가 증가하면 이러한 요구사항을 만족하기 위해 그에 비례하여 고성능 대용량 하드웨어가 필요하게 되며, 이는 하드웨어 인프라 비용 증가로 직결된다. 따라서, 하드웨어 인프라 비용을 줄이기 위해서는 기술의 복잡도를 줄이는 것이 필수적으로 요구되므로, 본 연구에서는 연산 복잡도가 상대적으로 낮은 1 단계 처리 방식의 객체 검출 기술인 YOLOv3 (J. Redmon, & Farhadi, A., 2018) 방식을 채택하였다.

3.2 문자열 인식

관심 문자열인 기기 고유번호 및 사용량 문자열 영역을 검출한 후, 검출된 영역 내에 존재하는 문자열을 인식해야 한다. 관심 문자열 영역을 정확하게 검출하였다면, 문자열 수만큼 영역을 등분할 한 후, 심층 신경망을 이용한 이미지 분류기를 이용하여 각 단위 문자를 인식하는 방법이 있다. 이러한 방법은, 기술 구조가 간단하다는 장점이 있으나, 문자열 영역이 정확하게 검출되지 않을 경우 인식 정확도에도 영향을 미칠 수 있다. 부정확한 문자열 영역 검출을 보완해 주기 위해서는 검출된 문자열 영역 내에서 단위 문자 영역을 다시 한 번 검출할 수도 있으나, 문자열 영역 검출에 이어 단위 문자를 별도로 검출해야 하고, 이미지 분류기를 단위 문자 개수만큼 처리해야 한다. 기기 고유번호는 12자리 단위 문자로 구성되어 있으며, 사용량은 4 자리 또는 5 자리의 단위 문자로 구성되어 있다. 따라서, 문자열 인식을 위해서는 총 16 또는 17 번의 심층 신경망 연산이 필요하므로, 전체 연산량이 과도하게 증가하는 단점이 있다.

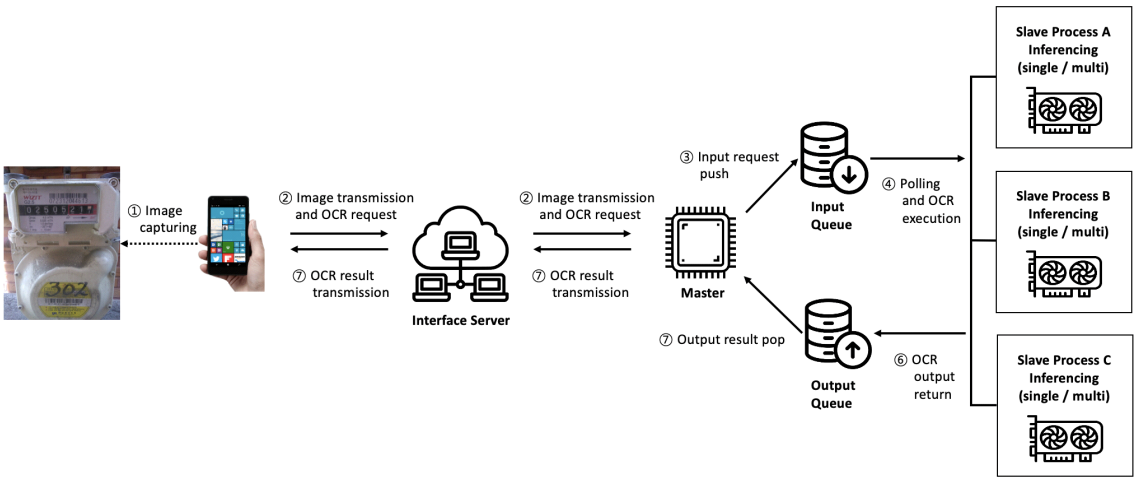
본 연구에서는 최소의 연산량으로 문자열을 인식하기 위해 공간-시계열 분석 구조를 가지는

CRNN (B. Shi et al., 2017) 신경망을 문자열 인식 구조로 채택하였다. 객체 검출에 의해 추정된 문자열 영역을 CRNN으로 입력하고, 합성곱 신경망을 적용하여 생성한 공간적 분석 특징 맵을 열 별로 시계열 분석한다. 시계열 분석을 통해 단위 문자가 존재하는 열은 해당 단위 문자로 매핑되고, 단위 문자가 존재하지 않은 열은 space 문자로 매핑한다. 최종 출력 시, space 문자가 나타나기 전까지 연속되는 동일 단위 문자 출력은 하나의 단위 문자로 변환하고, space 문자를 제거함으로써 최종 문자열을 구성한다.

CRNN 구조는 합성곱 신경망 및 순환 신경망으로 구성된 2 개의 신경망 연산만으로 문자열을 인식할 수 있으므로, 개별 단위 문자 인식보다 적은 연산량으로 전체 문자열을 인식할 수 있다.

3.3 자동 검침 시스템 구성

가스계량기 자동 검침을 위한 시스템은 크게 4가지 모듈로 구성되어 있다. 첫 번째는 모바일 기기를 이용하여 가스계량기 이미지를 획득하고 서버로 전송하기 위한 모바일 응용 소프트웨어이고, 두 번째는 모바일 기기에서 전송한 가스계량기 이미지를 수신하고 인식 결과를 모바일 기기로 회신하기 위한 인터페이스 서버이고, 세 번째는 인터페이스 서버로부터 전달받은 이미지에 대한 고속 병렬 처리를 위해 입출력 큐를 제어하는 마스터 소프트웨어이며, 마지막으로 네 번째는 관심 문자열 영역 검출 및 문자열 인식을 수행하는 심층 신경망 슬레이브 소프트웨어 이다. 전체적인 시스템 구성을 <Figure 2>에 도식화 하였다. 모바일 기기용 응용 소프트웨어는 계량기 사진을 획득하여, 기기 정보 및 접속 키를 인터페이스 서버로 송신하고, 회신 받은 분석 결과 정보를 시각화하여 사용자에게 전달한다. 모바일 기기와 인터페이스 서버 간의 통신은 가스 공급사가 사용 중인 Private LTE망을 이용하였다. 인터페이스 서버는 Private LTE망 기반에서 모바일 기기와 TCP/IP 프로토콜을 이용하여 이미지 및 분석 결과 정보를 주고받는다.



〈Figure 2〉 System Architecture of Automatic Gasometer Reading

GPU (Graphics Processing Unit)를 이용한 대용량 고속 병렬 처리를 위해 입출력 큐 구조를 채택하였다. 마스터 소프트웨어는 인터페이스 서버로 부터 전달받은 가스계량기 이미지 및 인식 요청 정보를 FIFO (First In First Out) 구조의 입력 큐에 넣는다. GPU에서 구동되는 심층 신경망 슬레이브 소프트웨어는 입력 큐의 상태를 상시 관찰하면서, 입력 큐에 인식 요청이 들어오면 관심 문자열 인식을 수행하고, 인식 결과 정보를 출력 큐에 넣는다. 출력 큐도 FIFO 구조를 채용하였다. 마스터 소프트웨어는 출력 큐를 상시 관찰하면서, 출력 큐에 인식 결과 정보가 들어오면 인식 결과 정보를 읽고, 인터페이스 서버를 통해 모바일 기기로 전송한다. 인식 결과 정보는 기물번호 문자열, 기물번호 영역 좌표, 사용량 문자열, 사용량 영역 좌표, 응답 코드 및 응답 메시지 등으로 구성되어 있다. 본 연구의 응용분야는 다수의 검침원이 전체 가스계량기 이미지를 동시 다발적으로 전송하여 문자열 인식을 요청하는 환경에 적용하였다. 문자열 인식 요청 빈도는 1일 약 70만 건에 이른다. 이러한 대용량 처리 요청에 대응하기 위해 병렬 처리가 가능한 복수개의 심층 신경망 슬레이브 프로세스를 구현하였다. 모든 슬레이브 프로세스는 IDLE 상태일 경우 입력 큐를 상시 관찰하면서 인식 요청을 기다린다. 인식 요청이 있을 경우, 가장 먼저 입력 큐를 접속한 슬레이브 프로세스는 BUSY 상태로 바뀌고, 심층 신경망을 구동하여 관심 문자열 인식을 수행한다. 관심 문자열 인식이 완료되면 출력 큐로 결과를 전달하고 다시 IDLE 상태로 전환되어 입력 큐 관찰을 시작한다.

전체 시스템은 Amazon Web Service 에서 제공하는 클라우드 환경에서 구현하였으며 인텔 제온 E5-2686 v4 CPU 및 Nvidia TESLA V100 GPU를 사용하였다. 심층 신경망 슬레이브 프로세스는 GPU에서 구동하였으며, 모바일 응용 소프트웨어는 모바일 기기에서 구동하였다. 그 외 모든 소프트웨어는 CPU에서 구동하였다.

4. 연구결과

4.1 데이터 및 레이블링

실험에 사용된 이미지 데이터는 일반 스마트폰을 이용하여 가스 사용자 거주지에 설치되어 실제 운영되고 있는 가스계량기를 촬영한 이미지를 사용하였다. 이미지 획득을 위한 별도의 모바일 어플리케이션을 구현하였으며, 촬영 시 가이드 박스를 표시하여 사용자가 계량기 촬영 시 사용량과 기기 고유번호를 가이드 박스 내에 크고 명확하게 위치하도록 조정하는 것을 유도하였다.

총 이미지 데이터는 27,120장으로 구성되어 있다. 이 중 학습용 (training)으로 18,389장을 사용하였고, 검증용 (validation)으로 4,596장을 사용하였으며, 최종 성능 테스트용으로 나머지 4,135장을 사용하였다. 신경망 학습 과정에 사용된 학습용 및 검증용 데이터는 총 22,985장으로써, 학습 차수 마다 무작위 혼합 방식을 적용하여 80%를 학습용으로, 20%를 검증용으로 할당함으로써 학습 차수 마다 서로 다른 조합의 학습용 데이터가 구성되도록 하였다.

테스트용 이미지는 기울어짐 없이 평행하고, 관심 문자열의 크기가 크고, 잡음이 첨가되지 않았고, 빛 반사가 없는 상태에서 촬영되어 육안으로 확인 시 뚜렷이 식별할 수 있는 normal 이미지 2,097장, 계량기가 이물질에 오염되었거나 촬

영 시의 손 떨림 등에 의해 이미지에 잡음 신호가 포함된 noise 이미지 655장, 계량기 일부분에 빛 반사가 포함된 reflex 이미지 653장, 멀리서 촬영하여 문자열 영역 크기가 작은 scale 이미지 268장 및 촬영 시 모바일 기기를 기울여서 찍음으로 인해 이미지가 기울어진 slant 이미지 462장으로 구성하였다. Normal 이미지를 제외한 나머지 분류의 이미지들은 일정 부분 왜곡이 포함되어 있는 비정상 이미지들로써, 왜곡 정도에 따른 인식률의 차이를 나타내기 위해 왜곡 정도를 level_1 및 level_2의 2단계로 상세 분류하였다. Level_1은 왜곡이 존재하나 정도가 심하지 않아서 육안으로 자세히 관찰 시 정보를 파악할 수 있어서 인식이 가능할 것으로 판단되는 것들이며, level_2는 왜곡의 정도가 상당히 심하여 육안으로 매우 자세히 관찰하여야만 문자열 확인이 가능한 이미지 데이터이다. 왜곡의 정도가 매우 심하여 육안으로 식별할 수 없는 이미지는 최종 데이터 셋에 포함시키지 않았다. Slant로 분류되는 데이터는 기울기가 5° ~ 10° 사이의 기울기를 가지는 이미지를 level_1로 상세 분류하고, 10° ~ 30° 사이의 기울기를 가지는 이미지를 level_2로 상세 분류하였다. 30° 이상의 기울기를 가지는 이미지는 수집한 데이터 셋 내에 존재하지 않았다. 기울어짐이 5° 이하이면 normal 이미지 데이터로 분류하였다. Scale로 분류되는 데이터는 문자열이 차지하는 영역이 전체 이미지 영역의 20% ~ 25%를 차지하면 level_1로 상세 분류하고, 영역이 차지하는 비율이 20% 이하이면 level_2로 상세 분류하였다. 사람의 눈으로 문자열 구별이 어려울 정도로 문자열 영역이 작은 이미지들은 최종 데이터 셋에서 제외하였다. Noise 및 reflex 로 분류되는 이미지들은 왜곡되지 않은 원본 이미지를 추정할 수 없으므로, 정량적인 기준에 의한 상세 왜곡 분류가 불가능하여, 개발자의 정성적인 기준으로 level_1 및 level_2로 상세 분류하였다. 개발자 1명이 단독으로 정성적 분류를 수행함으로써, 평가자에 따라서 왜곡 상세 분류에 차이가 날 수 있는 가능성을 제거하였다. Normal 데이터를 제외한 나머지 level_1, level_2를 통칭하여 abnormal 데이터라 칭한다.

객체 검출 기술을 이용하여 관심 문자열 영역을 검출하기 위해서는 데이터 레이블 생성 시, 관심 대상의 문자열 영역만을 레이블에 포함시켜야 한다. 본 연구에서의 관심 문자열은 가스계량기 고유 기물번호 문자열과 소수점을 제외한 사용량 문자열이다. 그 외의 나머지 문자열은 레이블에서 제외시킴으로써, 학습과정에서 관심 문자열과 비관심 문자열이 구분될 수 있도록 하였다.

〈Table 1〉 Total Data Amount

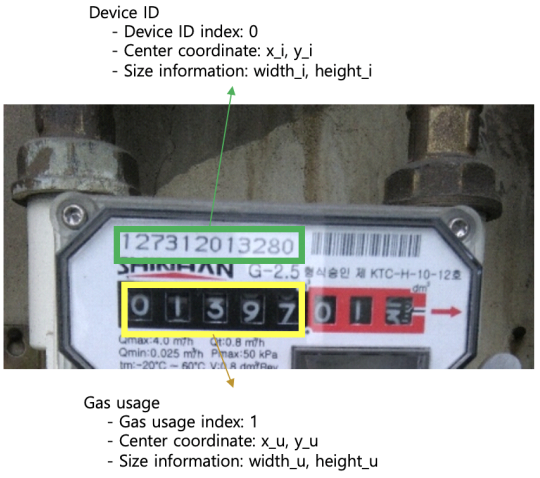
Category		Training and Validation	Test
Normal		22,985	2,097
Abnormal	Noise		655
	Reflex		653
	Scale		268
	Slant		462
총계		22,985	4,135

〈Table 2〉 Abnormal Test Data Details

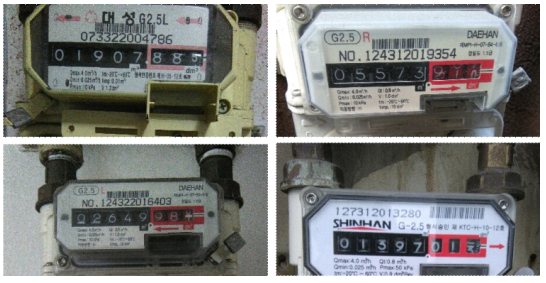
Abnormal	Level	
	Level_1	Level_2
Noise	486	169
Reflex	581	72
Scale	222	46
Slant	257	205

<Table 1>과 <Table 2>에 전체 데이터 및 왜곡 데이터 구성을 표기하였다.

레이블링은 가스계량기의 기기 고유번호 및 사용량 숫자 각각에 대해 문자열을 둘러싼 가장 작은 사각형의 위치 정보 및 사각형 내의 숫자열 정보를 레이블로 생성하였다. 또한 가스계량기 제조사 별로 기물번호, 사용량의 위치 및 형태가 다르므로, 관심 문자열 종류 정보를 별도 생성해야 한다. <Figure 3> 및 <Figure 4>에 레이블 데이터 생성 예시 및 제조사별 관심 문자열 배치 예시를 나타내었다.



〈Figure 3〉 Data Labeling Example



〈Figure 4〉 Various Character String Layout

4.2 실험 결과

계량기 기기 고유번호, 가스 사용량 두 가지 문자열에 대한 영역 검출 및 문자열 인식정확도 측정을 위해 IoU (Intersection over Union), 정밀도 (Precision), 재현율 (Recall) 및 문자열 인식 정확도 (Accuracy)를 기준으로 성능을 측정하였다. 문자열 영역 검출 성능 측정 지표인 IoU, 정밀도 및 재현율은 각 픽셀별로 검출된 영역이 레이블 영역 정보와 일치 하는지를 판별하여 측정하였으며, 문자열 인식 성능 측정 지표인 정확도는 추정된 문자열과 레이블 문자열의 일치 여부로 판별하였다. 문자열 영역 검출 및 인식에 대한 성능을 <Table 3>에 나타내었다. 문자열 인식 정확도 '전체'는 사용량 문자열과 기기 고유번호 문자열 인식 모두 오류가 없는 상태의 성능을 나타낸다.

<Table 3>은 객체 검출 성능 지표인 IoU, 정밀도, 재현율 및 최종 문자열 인식 정확도 성능 결과이다.

<Table 4>는 normal 이미지의 오류율 대비 abnormal 이미지 오류율의 비율을 배수로 표기한 것이다. <Table 4>에 표기된 오류율 증가 비율 Err.ratio는 (식 1) 과 같이 정의된다.

$$Err.ratio = \frac{1.0 - abnormal\ data\ accuracy}{1.0 - normal\ data\ accuracy} \quad (식1)$$

<Table 3>에서 보이듯이, normal, level_1 및 level_2 의 문자열 인식 평균 정확도는 기물번호 문자열이 각각 0.960, 0.870 및 0.562이며, 사용량 문자열은 각각 0.864, 0.757 및 0.601 이다. Normal 및 level_1의 문자열 인식 정확도는 기물번호가 사용량 보다 높으며, level_2의 문자열 인

〈Table 3〉 Performance of character region detection and character recognition

DB Category		DB size	Performance of character region detection						Character recognition accuracy		
			Usage			Device ID			Usage	Device ID	Total
			IoU	Recall	Prec.	IoU	Recall	Prec.			
Normal		2,097	0.744	0.981	0.756	0.803	0.972	0.823	0.864	0.960	0.835
Level_1	Noise	486	0.722	0.969	0.733	0.779	0.951	0.803	0.805	0.887	0.730
	Reflex	581	0.672	0.979	0.679	0.767	0.968	0.783	0.757	0.849	0.654
	Scale	222	0.691	0.983	0.701	0.771	0.956	0.799	0.743	0.851	0.635
	Slant	257	0.519	0.993	0.522	0.548	0.987	0.553	0.615	0.802	0.521
	Average		0.670	0.981	0.678	0.734	0.967	0.752	0.757	0.870	0.675
Level_2	Noise	169	0.662	0.982	0.669	0.736	0.947	0.768	0.746	0.515	0.402
	Reflex	72	0.683	0.963	0.694	0.737	0.951	0.753	0.625	0.556	0.306
	Scale	46	0.652	0.982	0.664	0.763	0.965	0.789	0.717	0.739	0.478
	Slant	205	0.353	0.985	0.354	0.348	0.975	0.350	0.161	0.132	0.024
	Average		0.604	0.979	0.612	0.664	0.961	0.682	0.601	0.562	0.377
Average			0.633	0.980	0.641	0.695	0.964	0.713	0.670	0.699	0.509

〈Table 4〉 Error rate increasement of abnormal data compared to normal data

DB Category		DB size	Performance ratio of character region detection						Character recognition accuracy ratio		
			Usage			Device ID			Usage	Device ID	Total
			IoU	Recall	Prec.	IoU	Recall	Prec.			
Level_1	Noise	486	1.09	1.63	1.09	1.12	1.75	1.11	1.43	2.83	1.64
	Reflex	581	1.28	1.11	1.32	1.18	1.14	1.23	1.79	3.78	2.10
	Scale	222	1.21	0.89	1.23	1.16	1.57	1.14	1.89	3.73	2.21
	Slant	257	1.88	0.37	1.96	2.29	0.46	2.53	2.83	4.95	2.90
Level_2	Noise	169	1.24	1.95	1.25	1.34	1.75	1.40	2.76	11.10	4.21
	Reflex	72	1.36	0.95	1.38	1.20	1.25	1.19	2.08	6.52	3.16
	Scale	46	2.53	0.79	2.65	3.31	0.89	3.67	6.17	21.70	5.92
	Slant	205	3.91	52.63	4.10	5.08	35.71	5.65	7.35	25.00	6.06

식 정확도는 기물번호가 사용량 보다 낮다. 이처럼 기물번호와 사용량 level_2 문자열 인식 정확도가 level_1과 다른 경향을 보이는 것은 normal 대비 기물번호 scale level_2 데이터의 IoU, 정밀도 오류율 증가가 level_1 오류율의 증가 보다 더 큰 것으로 부터 기인한다. 사용량 scale level_1 데이터의 normal 대비 IoU, 정밀도 오류율 증가는 각각 1.207배, 1.225배 이고, 기물번호 scale level_1 데이터의 normal 대비 IoU, 정밀도 오류율 증가는 각각 1.162배, 1.136배 로써, 사용량의 오류율 증가폭이 기물번호 대비 더 크다. 반면 사용량 scale level_2 데이터의 normal 대비 IoU, 정밀도 오류율 증가는 각각 2.527배, 2.648배 이고, 기물번호 scale level_2 데이터의 normal 대비 IoU, 정밀도 오류율 증가는 각각 3.310배, 3.672배 로써, 기물번호의 오류율 증가폭이 사용량 대비 더 크다. 기물번호 scale level_2 데이터의 문자열 인식 정확도 오류율의 normal 대비 증가폭도 21.7배로써 오류율 증가폭을 나타내는 전체 72개 성능 지표 중 가장 크다.

사용량 문자열 인식 오류율의 경우 normal 데이터 대비 abnormal level_1, level_2 데이터가 약 11% 및 26% 높은 오류율을 보이며, 기기 고유번호 문자열 인식 오류율의 경우 normal 데이터 대비 abnormal level_1, level_2 데이터가 각각 약 9% 및 40% 높은 오류율을 보이고 있다.

사용량과 기기 고유번호의 평균 문자열 인식률 차이는약 3%로써, 기기 고유번호가 사용량 대비 높은 인식률을 보인다. 반면, 객체 검출 재현율 측면에서는 사용량 재현율이 기기 고유번호 재현율대비 1.4% 높은 성능을 보이고 있다. 반대로, 객체 검출 IoU 및 정밀도는 기기 고유번호가 사용량 대비 각각 약 6.2%, 7.2% 높은 성능을 나타낸다. 따라서, 문자열 인식 정확도는 재현율 보다는 IoU 및 정밀도와 더욱 밀접한 비례 관계가 있음을 알 수 있다. 객체 검출 추정 영역이 실제 객체가 위치하고 있는 영역보다 넓은 영역으로 추정되면 재현율 성능 지표는 높게 나타나나, 실제 객체 영역 밖의 배경 이미지가 추정 영역에 포함되어 문자열 인식 모듈로 입력된다. 이 경우, 배경 이미지는 일종의 잡음으로 작용하여 문자열 인식 오류의 원인이 될 수 있다. 반면, IoU 및 정밀도는 추정 객체 영역과 실제 객체 영역이 유사한 경우에만 측정치가 높게 나타난다. 따라서, IoU 및 정밀도 성능 지표가 높을 경우 실제 문자열이 위치하고 있는영역 이외의 배경 이미지가 제거됨으로써, 문자열 인식 모듈로 입력되는 이미지에서 잡음 신호가 줄어들어 최종 문자열 인식 정확도가 높게 나타난다.

<Table 4>의 총 72개 성능 지표 중 abnormal 데이터의 오류율이 normal 데이터의 오류율 대비 감소하였음을 나타내는 1 이하의 지표는 총 5개로써 사용량 재현율 3개 및 기물번호 재현율 2개이다. 재현율의 오류율이 감소하였음에도 불구하고, 사용량 및 기물번호의 최종 문자열 인식 정확도는 normal 대비 모든 abnormal 데이터가 감소하는 것을 볼 수 있다. 기물번호 slant level_1 재현율의 경우 normal 대비 오류율이 약 50% 이상 감소하였음에도 불구하고 (기물번호 normal 재현율 오류율 0.028, 기물번호 slant level_1 재현율 오류율 0.013), 최종 문자열 인식 정확도 오류율은 normal 대비 slant level_1이 약 4.95배 증가한다. 기물번호 slant level_1 데이터 이 외에 재현율의 오류율이 감소한 나머지 4개 abnormal 데이터도, 재현율의 오류율이 감소하였으나, 최종 문자열 인식 정확도는 증가하는 것을 볼 수 있다.

<Table 4>에서 사용량과 기기 고유번호의

Level_2 Slant 항목 재현율 성능 하락 정도는 각각 52.63, 32.71배로써 다른 항목 대비 매우 크게 나타난다. Level_2 Slant 데이터는 기울어진 정도가 10° ~ 30° 로써, 다른 종류의 왜곡과는 다르게 이미지가 많이 기울어질수록 문자열 영역 이 외의 잡음 영역이 객체 추정 영역에 포함된다. 이러한 이유로, Level_2 Slant 데이터의 경우, 찾고자 하는 정확한 문자열 영역 이 외의 잡음 영역이 일정 비율 이상으로 증가하면서 객체 검출 심층 신경망이 잡음 영역을 객체 모델에 반영하게 되어 관심 문자열 영역 검출에 부정적인 영향을 준 것으로 추정된다. 이러한 문제점은 객체 분할 또는 좌표 추정과 같은 기술을 이용하여 잡음 영역을 제거하면 개선될 것으로 기대한다.

<Table 3>를 살펴보면, 전반적으로 scale 및 slant 데이터에서 문자열 인식 정확도가 떨어지는 것을 볼 수 있다. Scale level_1 데이터의 경우 사용량과 기물번호의 평균 문자열 인식정확도 오류율이 normal 데이터 대비 약 2.8배 증가하며, slant level_1의 경우 약 3.9배 증가한다. Scale level_2 및 slant level_2의 사용량과 기물번호의 평균 문자열 인식 정확도 오류율은 각각 13.9배 및 5.0배 증가한다. 문자열 인식 정확도 오류율의 증가 폭이 가장 큰 데이터는 기물번호 scale level_2 로써, normal 데이터 대비 약 21.7배 까지 증가한다. Abnormal 상세 분류별 인식률을 살펴보면, noise 및 reflex 데이터 보다는 scale 및 slant 데이터에서 인식률이 낮게 나타난다. 특히 slant 데이터의 인식률 저하가 낮게 나타나는데, level_1 slant 데이터의 경우 사용량과 기물번호 문자열 인식률은 각각 0.615, 0.802로써 normal 대비 0.249, 0.158 하락하나, level_2 slant 데이터의 경우 사용량와 기물번호 문자열 인식률은각각 0.161, 0.132로써 normal 대비 성능 하락 정도가 각각 0.596, 0.738 이다.

5. 결론

본 연구에서는 가스계량기 사진을 분석하여 가스 사용량 및 계량기 고유번호를 인식하고 가스 사용량에 대한 과금을 자동 처리할 수 있는 응용 시스템을 구축하였다. 이를 위해 이미지 내의 많은 문자열들 중 관심 대상 문자열만을 검출하여 인식할 수 있는 심층 신경망 기술 구조를 제시하였다. AWS (Amazon Web Service) 클라우드 플랫폼에서 제공하는 인텔 제온 E5-2686 v4 CPU 및 Nvidia TESLA V100 GPU를 이용하여 전체 시스템을 구축하였으며, 1일 처리 가능한 자동 검침 요청은 약 70만 건이다.

관심 문자열 인식을 위한 심층 신경망 구조는 관심대상인 기물번호 및 사용량 문자열 위치를 검출하기 위한 객체 검출 심층 신경망과 검출된 영역 내의 분자열을 인식하는 문자열 인식 심층 신경망으로 구성되어 있다. 12자리로 구성된 기물번호 문자열 인식률이 4 ~ 5자리로 구성된 사용량 문자열 인식률 보다 더 높다. 일반적으로는 문자열의 자리수가 작을수록 인식률은 좋아지나, 본 연구 결과에서 보이듯이 영역 검출 IoU 및 정밀도가 낮으면 문자열 자리수가 작아도 오류율이 높을 수 있다. 이는 일정 부분 CRNN 신경망의 구조적 특성과도 관계가 있을 것으로 추정된다. CRNN 신경망은 공간적 분석과 시계열 분석을 순차적으로 처리하는 구조로써, 공간적 분석 신경망의 분석 결과가 시계열 분석 신경망으로 입력되므로, IoU 및 정밀도가 낮은 영역 검출 결과로 인해 시계열 분석 신경망의 입력 데이터의 공간적 일관성이 사라지면서 문자열 인식

결과가 낮게 나타났다고 볼 수 있다. 즉, CRNN으로 입력되는 데이터에 문자열 영역 이 외의 다른 영역이 많이 포함될수록 인식률이 저하되는 경향이 있는데, 이는 CRNN이 배경 이미지를 space 문자로 매핑하지 못한 것으로 보인다. 문자열 영역 검출에 오류가 있을 경우, 다양한 배경 이미지가 CRNN 으로 입력되며, CRNN은 이러한 모든 다양한 배경 이미지를 동일한 space 문자로 할당해야 하는데, 배경 이미지 간 데이터 패턴이 서로 다를 경우 이러한 매핑에 문제가 발생할 수 있다.

CRNN의 배경 이미지 처리 문제는 slant 데이터의 문자열 인식률 하락에서도 나타난다. 문자열이 기울어진 slant 데이터의 경우, 문자열 영역 내의 아래 위 공간에 배경 이미지가 포함된다. 이러한 문자열 영역 내의 배경 이미지도 일종의 잡음으로 작동할 수 있다. Slant 데이터는 기울어진 정도에 따라 잡음의 양이 달라질 수 있다. Slant level_2 데이터의 normal 대비 성능 하락 정도가 slant level_1 데이터의 normal 대비 성능 하락 정도보다 매우 큰 것도 CRNN의 배경 이미지 처리 문제로 볼 수 있다.

6. 추가연구

인식하고자 하는 관심 대상 문자열이 가로 또는 세로형태의 직사각형이 아닌 기울어진 형태 등 직사각형 이 외의 형태로 배치되어 있을 경우 관심 대상 문자열 영역 이 외에 배경 이미지 정보들이 문자열인식 대상 영역에 포함되므로, 관심 문자열 검출 이 후에 적용될 문자열 인식 기술의 성능 저하 원인이 될 수 있다. 관심 대상 문자열의 영역이 정사각형이 아닌 임의 형태로 배치되어있거나, 잡음 및 빛 반사가 있는 경우에도 관심 문자열 영역을 정확하게 검출할 수 있어야 한다. 이러한 문제점을 보완하기 위해 2 가지 추가 연구가 필요할 것으로 판단된다.

첫째는 관심 문자열 영역을 보다 정확하게 추정하기위해 자세 추정에 사용되는 기술들을 접목하는 것이다. 자세 추정 기술은 사람의 관절 포인트 위치를 추정하는 기술로써, 관심 문자열 영역의 4개 꼭짓점 위치를 직접 추정하는 방법을 시도해 볼 수 있다. 자세 추정에도 심층 신경망을 적용할 수 있는데, 두 개의 심층 신경망을 순차 연결하여 관절 포인트 위치 추정오차를 보정하는 방법 (Toshev& Szegedy, 2014), 관절 포인트 좌표와 주변 일정 영역 내 픽셀들의 상대적 위치차이를 모델링하는 방법 (Papandreouet al., 2017), 서로 다른 관절 포인트 위치를 동시에 모델링하기 위해 분석 영역을 점진적으로 확대해 나가는 방법 (Wei, Ramakrishna, Kanade, & Sheikh, 2016), 서로 다른 관절 포인트의 상대적 위치 벡터를 직접 모델링하는 방법 (Cao,Simon, Wei, & Sheikh, 2017)등이 있다. 이러한 자세 추정 방법을 적용하여 직접 관심 문자열 영역 꼭짓점 좌표를 추정할 수도 있고, 기존의 객체 검출 결과에 적용하여 2 단계 기술 구조를 통해 점진적으로 영역 위치 오류를 제거할 수도 있을 것이다.

둘째는 객체 분할 기술을 적용하여 관심 문자열 영역을 추정하는 것이다. 객체 분할 기술은 객체가 차지하는 영역의 모양을 사전에 정의하지 않고, 각 픽셀이 객체 영역 내에 포함되는지 여부를 판별하는 기술이다. 객체 분할 기술로는 인코더와 디코더 구조를 다르게 함으로써 적은 연산량으로 영역 분할이 가능한 방법 (Paszke, 2016), 합성곱 신경망으로만 영역 분할 신경망을 구성하는 방법 (Long, 2015), 객체 검출 기술인

R-CNN 신경망을 변형하여 적용하는 방법 (K.He, Gkioxari, G., Dollár, P., & Girshick, R., 2017)등이 있다. 객체 분할 기술 역시 관심 문자열 영역을 직접 추정하는 방법으로 적용할 수도 있고, 기존의 객체 검출 결과에 적용하여 2 단계 기술 구조를 통해 점진적으로 영역 위치 오류를 제거할 수도 있을 것이다. 객체 영역 검출 정확도가 높아지면 CRNN의 배경 이미지 입력 문제도 완화시킬 수 있을 것으로 기대한다.

참고문헌(References)

Ahn, H., K.-j. Kim, and I. Han, "Purchase Prediction Model using the Support Vector Machine," *Journal of Intelligence and Information Systems*, Vol.11, No.3(2005), 69～81.

Baek, Y., Lee, B., Han,D., Yun, S., & Lee, H. (2019). Character Region Awareness for TextDetection. arXiv preprintarXiv:1904.01941.

Ballard, D. H. (1981).Generalizing the Hough transform to detect arbitrary shapes. Pattern recognition, 13(2), 111-122.

Canziani, A., Paszke, A., & Culurciello, E. (2016). An analysis of deep neural network models forpractical applications. arXiv preprintarXiv: 1605.07678.

Cao, Z., Simon, T., Wei,S.-E., & Sheikh, Y. (2017). Realtimemulti-person 2d pose estimation using part affinity fields. Paper presentedat the Proceedings of the IEEE conference on computer vision and patternrecognition.

Dalal, N., & Triggs,B. (2005). Histograms of orientedgradients for human detection. Paper presented at the international Conferenceon computer vision & Pattern Recognition (CVPR'05).

Felzenszwalb, P. F., &Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. International journal of computer vision, 59(2), 167-181.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of patternrecognition unaffected by shift in position. Biological cybernetics, 36(4), 193-202.

Girshick, R., Donahue, J.,Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurateobject detection and semantic segmentation. nProceedings of the IEEE conference on computer vision and pattern recognition(pp. 580-587).

Girshick, R., Donahue, J.,Darrell, T., & Malik, J. (2015). Fast r-cnn. n Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

Glorot, X., & Bengio,Y. (2010). Understanding the difficultyof training deep feedforward neural networks. Paper presented at theProceedings of the thirteenth international conference on artificialintelligence and statistics.

Gower, J. C., & Ross,G. J. (1969). Minimum spanning trees and single linkage cluster analysis. Applied statistics, 54-64.

He, K., Gkioxari, G.,Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision(pp. 2961-2969).

He, K., Zhang, X., Ren,S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutionalnetworks for visual recognition.

IEEEtransactions on pattern analysis and machine intelligence, 37(9), 1904-1916.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conferenceon computer vision and pattern recognition.

Hochreiter, S., &Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Huang, G., Liu, Z., VanDer Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Paper presented at the Proceedingsof the IEEE conference on computer vision and pattern recognition.

Huang, Z., Xu, W., &Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991.

Iandola, F., Moskewicz,M., Karayev, S., Girshick, R., Darrell, T., & Keutzer, K. (2014). Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv: 1404.1869.

Ioffe, S., & Szegedy,C. (2015). Batch normalization: Accelerating deep network training by reducinginternal covariate shift. In Internationalconference on machine learning.

Jaderberg, M., et al.(2016). Reading Text in the Wild with Convolutional Neural Networks. International Journal of Computer Vision,vol. 116, no. 1, 2016, pp. 1-20.

Krizhevsky, A., Sutskever,I., & Hinton, G. E. (2012). Imagenet classification with deep convolutionalneural networks. In Advances in neuralinformation processing systems (pp. 1097-1105).

Larsson, G., Maire, M., & Shakhnarovich, G. (2016). Fractalnet: Ultra-deep neural networks withoutresiduals. arXiv preprintarXiv:1605.07 648.

LeCun, Y., Bottou, L.,Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied todocument recognition. Proceedings of theIEEE, 86(11), 2278-2324.

Liao, M., Shi, B., &Bai, X. (2018). Textboxes++: A single-shot oriented scene text detector. IEEE transactions on image processing, 27(8), 3676-3690.

Liao, M., Shi, B., Bai, X., Wang, X., & Liu, W. (2017). Textboxes:A fast text detector with a single deep neural network. Paper presented atthe Thirty-First AAAI Conference on Artificial Intelligence.

Lin, M., Chen, Q., & Yan, S. (2013). Network in network. arXivpreprint arXiv:1312.4400.

Lindeberg, T. (2012). Scale invariant feature transform.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector.Paper presented at the European conference on computer vision.

Liu, W., Chen, C., Wong, K.-Y. K., Su, Z., & Han, J. (2016). STAR-Net:A SpaTial Attention Residue Network for Scene Text Recognition. Paperpresented at the BMVC.

Long, J., Shelhamer, E.,& Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference oncomputer vision and pattern recognition (pp. 3431-3440).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International

journal of computer vision, 60(2), 91-110.

Nair, V., & Hinton, G.E. (2010). Rectified linear units improverestricted boltzmann machines. Paper presented at the Proceedings of the27th international conference on machine learning (ICML-10).

Papandreou, G., Zhu, T.,Kanazawa, N., Toshev, A., Tompson, J., Bregler, C., & Murphy, K. (2017). Towards accurate multi-person poseestimation in the wild. Paper presented at the Proceedings of the IEEEconference on computer vision and pattern recognition.

Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). Enet: A deep neural network architecture for real-time semantic segmentation. arXivpreprint arXiv:1606.02147.

Redmon, J., & Farhadi,A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and patternrecognition (pp. 7263-7271).

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Redmon, J., Divvala, S.,Girshick, R., & Farhadi, A. (2016). Youonly look once: Unified, real-time object detection. Paper presented at theProceedings of the IEEE conference on computer vision and pattern recognition.

Ren, S., He, K., Girshick,R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection withregion proposal networks. In Advances inneural information processing systems (pp. 91-99).

Rosenblatt, F. (1958). Theperceptron: a probabilistic model for information storage and organization inthe brain. Psychological review, 65(6), 386.

Rumelhart, D. E., Hinton,G. E., & Williams, R. J. (1986). Learning representations byback-propagating errors. nature, 323(6088), 533-536.

Shi, B., Bai, X., &Yao, C. (2017). An end-to-end trainable neural network for image-based sequencerecognition and its application to scene text recognition. IEEE transactions on pattern analysis and machine intelligence, 39(11), 2298-2304.

Shi, B., et al. (2017).Detecting Oriented Text in Natural Images by Linking Segments. 2017 IEEE Conference on Computer Vision andPattern Recognition (CVPR), 2017, pp. 3482-3490.

Simonyan, K., & Zisserman,A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Szegedy, C., Ioffe, S.,Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet andthe impact of residual connections on learning. In Thirty-First AAAI Conference on Artificial Intelligence.

Szegedy, C., Liu, W., Jia,Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition(pp. 1-9).

Tian, S., Lu, S., &Li, C. (2017). Wetext: Scene text detection under weak supervision. Paper presented at the Proceedings of theIEEE International Conference on Computer Vision.

Toshev, A., & Szegedy, C. (2014). Deeppose: Human poseestimation via deep neural networks. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Uijlings, J. R., Van DeSande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search forobject recognition. International journalof computer vision, 104(2), 154-171.

Wei, S.-E., Ramakrishna,V., Kanade, T., & Sheikh, Y. (2016). Convolutionalpose machines. Paper presented at the Proceedings of the IEEE conference oncomputer vision and pattern recognition.

Zeiler, M. D., &Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. European Conference on Computer Vision, 818-833.

Zhou, X., Yao, C., Wen,H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). EAST: an efficient and accurate scene text detector. Paper presented at the Proceedings of the IEEE conference on computer vision andpattern recognition.

Zhu, X., etal. (2017). Deep Residual Text Detection Network for Scene Text. 2017 14th IAPR International Conference onDocument Analysis and Recognition (ICDAR), 2017, pp. 807-812.

Abstract

Automatic gasometer reading system using selective optical character recognition

Kyohyuk Lee* · Taeyeon Kim** · Wooju Kim***

In this paper, we suggest an application system architecture which provides accurate, fast and efficient automatic gasometer reading function. The system captures gasometer image using mobile device camera, transmits the image to a cloud server on top of private LTE network, and analyzes the image to extract character information of device ID and gas usage amount by selective optical character recognition based on deep learning technology. In general, there are many types of character in an image and optical character recognition technology extracts all character information in an image. But some applications need to ignore non-of-interest types of character and only have to focus on some specific types of characters. For an example of the application, automatic gasometer reading system only need to extract device ID and gas usage amount character information from gasometer images to send bill to users. Non-of-interest character strings, such as device type, manufacturer, manufacturing date, specification and etc., are not valuable information to the application. Thus, the application have to analyze point of interest region and specific types of characters to extract valuable information only. We adopted CNN (Convolutional Neural Network) based object detection and CRNN (Convolutional Recurrent Neural Network) technology for selective optical character recognition which only analyze point of interest region for selective character information extraction. We build up 3 neural networks for the application system. The first is a convolutional neural network which detects point of interest region of gas usage amount and device ID information character strings, the second is another convolutional neural network which transforms spatial information of point of interest region to spatial sequential feature vectors, and the third is bi-directional long short term memory network which converts spatial sequential information to character strings using time-series analysis mapping from feature vectors to character strings. In this research, point of interest

* Management of Technology, Yonsei University
** Computer Science, KAIST
*** Corresponding Author: Wooju Kim
Graduate School of Information and Industrial Engineering, Yonsei University
50 Yonsei-ro Seodaemoon-gu, Seoul, Republic of Korea
Tel: +82-2-2123-5716, E-mail: wkim@yonsei.ac.kr

character strings are device ID and gas usage amount. Device ID consists of 12 arabic character strings and gas usage amount consists of 4 ~ 5 arabic character strings. All system components are implemented in Amazon Web Service Cloud with Intel Zeon E5-2686 v4 CPU and NVidia TESLA V100 GPU. The system architecture adopts master-lave processing structure for efficient and fast parallel processing coping with about 700,000 requests per day. Mobile device captures gasometer image and transmits to master process in AWS cloud. Master process runs on Intel Zeon CPU and pushes reading request from mobile device to an input queue with FIFO (First In First Out) structure. Slave process consists of 3 types of deep neural networks which conduct character recognition process and runs on NVidia GPU module. Slave process is always polling the input queue to get recognition request. If there are some requests from master process in the input queue, slave process converts the image in the input queue to device ID character string, gas usage amount character string and position information of the strings, returns the information to output queue, and switch to idle mode to poll the input queue. Master process gets final information form the output queue and delivers the information to the mobile device. We used total 27,120 gasometer images for training, validation and testing of 3 types of deep neural network. 22,985 images were used for training and validation, 4,135 images were used for testing. We randomly splitted 22,985 images with 8:2 ratio for training and validation respectively for each training epoch. 4,135 test image were categorized into 5 types (Normal, noise, reflex, scale and slant). Normal data is clean image data, noise means image with noise signal, relfex means image with light reflection in gasometer region, scale means images with small object size due to long-distance capturing and slant means images which is not horizontally flat. Final character string recognition accuracies for device ID and gas usage amount of normal data are 0.960 and 0.864 respectively.

Key Words : Gasometer, automatic reading, selective optical character recognition, convolutional neural network, recurrent neural network, parallel processing

Received : April 11, 2020 Revised : May 11, 2020 Accepted : May 14, 2020
Publication Type : Regular Paper Corresponding Author : Wooju Kim

저 자 소 개

이 교 혁

삼성전자 DMC 연구소, SK 하이닉스 멀티미디어 연구소에서 인공지능 영상 분석, 음성 인식, 비디오 코덱 등을 연구하였으며, 현재 ㈜카이어 대표이사로 재직 중이다. 연세대학교 공학박사 과정에 재학 중이며, 관심 연구 분야는 인공지능 컴퓨터 비전, 음성 인식, 이상 탐지 및 빅데이터 분석 등이다.

김 태 연

KAIST 전산학부 컴퓨터공학 과정에서 2017년 학사 학위 및 2019년 석사 학위를 취득하고, 인공지능 영상처리 스타트업에서 근무 중이다. 관심 연구 분야는 컴퓨터 비전, 문자 인식 및 디지털 포렌식 등이다. 딥러닝 연구자로서 다수의 컴퓨터 비전 프로젝트에 참여하였다

김 우 주

1987년 연세대학교 BBA 과정 학사 학위를 취득하고, 1994년 KAIST 경영과학 박사를 취득하였으며, 현재 연세대학교 정보산업공학과 교수로 재직 중이다. 관심분야는 시맨틱 웹, 시맨틱 웹 환경의 의사결정지원 시스템, 시맨틱 웹 마이닝, 지식관리 및 인공지능 웹 서비스이다.