



저작자표시 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#) 

공학석사 학위논문

아날로그 게이지 인식 및 정보 획득

Analogue Gauge Recognition and Analysis of Information

울 산 대 학 교 대 학 원
전기전자컴퓨터공학과
손 기 석

아날로그 게이지 인식 및 정보 획득

지도교수 조강현

이 논문을 공학석사학위 논문으로 제출함

2019 년 7 월

울산대학교 대학원
전기전자컴퓨터공학과
손기석

손기석의 공학석사학위 논문을 인준함

심사위원장 강 희 준



심 사 위 원 조 강 현



심 사 위 원 서 영 수



울 산 대 학 교 대 학 원

2019 년 8 월

감사의 글

대학원 석사과정을 진행하면서 저의 학위논문이 잘 마무리될 수 있었던 것은 많은 분들의 도움이 있었기에 가능하였습니다. 이 글을 통해 감사의 인사를 드리고자 합니다.

먼저, 석사과정 동안 부족한 저를 이끌어 주시고 지원해주신 지도교수 조강현 교수님께 감사드립니다. 그리고 더 좋은 논문이 될 수 있도록 조언과 격려를 해주신 강희준 교수님, 서영수 교수님께 감사의 말씀을 전합니다.

연구실에서 항상 어려운 일이 있을 때마다 도와주시고 힘이 되어주신 율경이형 고맙습니다. 그리고 연구를 위해 새로운 아이디어를 제공해주신 진수와 연구실 멤버 모두에게 고맙다는 말을 전하고 싶습니다.

먼저 연구를 진행하여 나를 이끌어준 범석이와 마지막까지 연구를 같이 진행하며 많이 도와준 재완아 고맙다.

지면으로 미처 언급하지 못했지만, 저를 아끼고 격려해 주셨던 모든 분들께도 감사하다는 말씀을 드립니다. 더욱 빛나도록 앞으로 최선을 다하겠습니다.

마지막으로 항상 저를 믿고 지원해주신 우리가족들 감사합니다.

감사합니다.

[국문 요약]

아날로그 게이지 인식 및 정보 획득

울산대학교 대학원
전기전자컴퓨터공학과
손 기 석

글로벌 금융 위기 이후 제조업은 장기전인 경기 침체를 겪었다. 이를 타개 하기 위해 정부는 대규모 투자와 정책을 수립하였으며 최근 4차 산업혁명으로 대두된 스마트 팩토리를 구축하기 시작했다. 스마트 팩토리는 ICT기술이 융합되어 있으며 각종 센서와 장치, 기계와 연동되어 효율적인 제품 생산을 돕는다. 아날로그 게이지는 공장에서 사용되는 필수 계측장비로 많은 공장에서 사용되나 안전불감증과 비용의 문제로 투자대상에서 제외되고 있다. 기존의 아날로그 게이지를 디지털화한다면 인적, 물적 피해를 방지할 수 있으며 스마트 팩토리의 구축에 비용절감으로 이어진다.

본 논문에서는 아날로그 게이지의 정보를 획득하는 두 가지의 알고리즘을 제시하며 첫 번째는 영상처리를 이용한 알고리즘이며 두 번째는 Fully Convolutional Network(FCN)를 이용하는 알고리즘이다. 영상처리를 이용한 알고리즘은 다양한 환경에 적합하지 못하여 결과가 좋지 못하였으나 FCN을 이용한 알고리즘은 Tesseract-OCR을 단독으로 사용하였을 때 보다 F1 score기준 2.1배 더 좋은 결과를 보여 줬다.

[영문 요약]

Analogue Gauge Recognition and Analysis of Information

Gi-Seok Sohn

School of Electrical Engineering,

The Graduate School,

University of Ulsan

Supervised by Prof. Kang-Hyun Jo

ABSTRACT

Since the global financial crisis, manufacturing has suffered from a long-term recession. To solve this problem, the government has established a large-scale investment and policy, and recently began building smart factories that show up as the fourth industrial revolution.. The Smart Factory is a fusion of ICT technology and works efficiently with various sensors, devices and machines. Analog gauges are essential instrumentation equipment used in factories and are used in many factories, but they are excluded from the investment due to safety insufficiency and cost. Digitizing existing analog gauges can prevent human and material damage and leads to cost savings in building smart factories.

In this paper, we propose two algorithms for acquiring the information of analog gauge. The first is algorithm using image processing and the second is algorithm using Fully Convolutional Network (FCN). The algorithm using image processing was not suitable for various environments and the results were not good. However, the algorithm using FCN showed 2.1 times better than the Tesseract-OCR by F1 score alone.

목 차

국문요약	I
영문요약	II
목 차	III
그림목차	V
표 목 차	VII
1. 서 론	1
1.1. 연구의 배경 및 목적	1
1.2. 논문 구성	3
2. 연구의 이론적 배경	4
2.1. 디지털 영상처리	4
2.1.1. 이진화 임계 분할(Binarization Threshold)	4
2.1.2. 모폴로지(Morphology)	5
2.1.3. 컨투어(Contour)	8
2.1.4. 허프 (Hough Transform)	8
2.2. 심층학습(Deep Learning)	12
2.2.1. 이미지 인식(Image Recognition)	12
2.2.2. 퍼셉트론(Perceptron)	15
2.2.3. 활성화 함수(Activation Function)	17
2.2.4. 컨볼루션 신경망(Convolutional Neural Network, CNN)	19
2.2.5. FCN(Fully Convolutional Network)	22
2.2.6. 최소자승법	24
2.3. Tesseract-OCR	25
3. 아날로그 페이지 정보 획득	26
3.1. 영상처리를 이용한 아날로그 페이지 정보 획득	26
3.2. FCN을 이용한 아날로그 페이지 정보 획득	32
4. 실험	38
4.1. 실험환경	38
4.2. 영상처리를 이용한 아날로그 페이지 정보 획득 실험	39

4.3. FCN을 이용한 아날로그 게이지 정보 획득 실험	41
5. 결 론	45
참 고 문 헌	46

그림 목 차

그림 1.1	전세계 스마트 팩토리 시장 규모	1
그림 2.1	원본이미지(좌)와 이진화 임계 분할 결과(우)	4
그림 2.2	침식 연산의 예시	5
그림 2.3	팽창 연산의 예시	6
그림 2.4	열림 연산의 예시	7
그림 2.5	단힘 연산의 예시	7
그림 2.6	원본이미지(좌)와 컨투어 적용 결과(우)	8
그림 2.7	직선의 구조	8
그림 2.8	xy평면과 ab평면의 관계	9
그림 2.9	직선을 r과 θ 의 변환	10
그림 2.10	xy평면과 $\rho\theta$ 평면의 관계	10
그림 2.11	원의 구조	11
그림 2.12	원에 대한 허프 변환	12
그림 2.13	나무에 사과가 있는 이미지	13
그림 2.14	심층 학습을 이용한 이미지 분류의 예시	14
그림 2.15	심층 학습을 이용한 이미지 탐지의 예시	14
그림 2.16	심층 학습을 이용한 이미지 시맨틱 세그멘테이션의 예시	14
그림 2.17	신경세포와 단일 퍼셉트론	15
그림 2.18	다층 퍼셉트론	16
그림 2.19	계단 함수	17
그림 2.20	시그모이드 함수	18
그림 2.21	ReLU 함수	18
그림 2.22	컨볼루션 신경망의 구조	19
그림 2.23	컨볼루션 연산	20
그림 2.24	최대풀링 과정(좌), 최대풀링 결과(우)	21
그림 2.25	CNN과 FCN의 차이	22
그림 2.26	FCN의 구조	23
그림 2.27	FCN의 업샘플링 방법	23
그림 2.28	최소자승법의 원리	24
그림 2.28	Tesseract-OCR의 적용 예시	25
그림 3.1	영상처리를 이용한 아날로그 게이지 정보 획득	26
그림 3.2	아날로그 게이지 이미지(좌)와 전처리의 결과(우)	27
그림 3.3	원에 대한 허프 변환의 결과	27
그림 3.4	방사 탐색(좌)과 방사 탐색 결과(우)	28

그림 3.5	덩어리를 포함하는 가장 작은 원	28
그림 3.6	각도에 다른 덩어리의 수 히스토그램	29
그림 3.7	OCR 결과(좌)과 획득한 정보(우)	29
그림 3.8	동심원에 다른 덩어리의 수 히스토그램	30
그림 3.9	관심 영역 설정 결과	30
그림 3.10	OCR 결과(좌)과 획득한 정보(우)	31
그림 3.11	비율 계산(좌)과 각도 계산(우)	31
그림 3.12	FCN을 이용한 아날로그 페이지 정보 획득	32
그림 3.13	ICDAR 데이터의 구성	34
그림 3.14	아날로그 페이지 데이터의 구성	35
그림 3.15	Tesseract-OCR의 판독 결과	36
그림 3.16	최소자승법을 이용한 중심 추출	37

표 목 차

표 4.1	영상처리를 이용한 아날로그 페이지 정보 획득 실험 결과	39
표 4.2	알고리즘의 추가 분석	40
표 4.3	Tesseract-OCR만 사용한 경우에 대한 결과	41
표 4.4	FCN을 이용한 아날로그 페이지 정보 획득	42
표 4.5	전체 테스트 데이터 셋에 대한 결과	43
표 4.6	FCN을 이용한 아날로그 페이지 정보 획득 실험 결과	44

1. 서론

1.1 연구의 배경 및 목적

글로벌 금융위기 이후 세계 제조업은 장기적인 경기침체와 노동 원가 및 원자재 비용 상승, 늦은 공정화 혁신 등으로 성장 한계에 봉착했다. 전 세계는 제조업의 중요성을 깨닫고 난관을 타개하기 위해 대규모 투자와 정책수립을 하였으며 4차 산업혁명으로 대두된 스마트 팩토리 구축하기 시작했다. 스마트 팩토리는 공장의 모든 센서나 기계의 장치가 ICT기술과 융합되어 제품 생산의 복잡한 모든 과정을 하나의 공장이 움직이는 것처럼 상호 연동되어 움직이는 공장을 뜻한다. 이러한 시스템은 기존의 대량생산을 위한 자동화 공장에서 벗어나 소비자가 원하는 다양한 제품을 효율적으로 생산할 수 있다. 스마트 팩토리를 구성하는 요소는 IoT, AI, 빅데이터 등이 있다.



그림1.1 전세계 스마트 팩토리 시장 규모[1]

아날로그 게이지는 공장에서 많이 사용되는 계측 장비로 최근에는 디지털 게이지로 변환되고 있다. 기존의 아날로그 게이지를 디지털 장치로 변환하면 사람에 의한 인적, 물적 재해 발생 방지를 할 수 있으나 이를 위해 소요되는 인적, 경제적 비용이 높으며 특히 산업체 안전 불감증에 의해 투자 대상에서 제외되고 있다. 기존 아날로그 게이지를 디지털화할 수 있는 장비를 개발하면 별도의 교체 없이 인적, 물

적 재해 발생 방지를 할 수 있고 스마트 팩토리와 연계해서 사용할 수 있으므로 아날로그 게이지 인식 시스템 개발이 필요하다.

본 논문에서는 아날로그 게이지의 정보 획득을 위한 방법을 제시한다. 제시한 방법은 두 가지이며 첫 번째는 영상처리의 기술을 적용한 방법이며 영상처리의 기술을 적용한 방법의 문제점을 해결하기 위해 고안한 두 번째 방법은 Fully Convolutional Network(FCN)[2, 3]를 적용한 방법이다.

첫 번째는 영상처리의 기술을 이용한 방법이다. 아날로그 게이지 이미지를 영상 분할하여 이진화시키고 노이즈 제거와 결손 된 부분을 보완하기 위해 모폴로지(Morphology) 방법을 사용한다. 이후 원에 대한 허프 변환을 적용하여 아날로그 게이지의 중심점을 찾는다. 아날로그 게이지의 지침을 찾기 위해 아날로그 게이지의 중심점으로부터 방사 탐색을 한다. 여기서 방사 탐색을 일정한 간격의 동심원의 형식으로 탐지함을 의미한다. 방사 탐색을 얻은 결과를 토대로 인접한 픽셀의 컨투어(Contour)를 만들고 컨투어를 포함하는 가장 작은 원을 그린다. 가장 작은 원의 중심을 찾고 중심으로부터 각 작은 원의 중심의 각도를 계산하고 각도가 가장 많이 겹치는 작은 원이 아날로그 게이지의 지침이 된다. 아날로그 게이지의 숫자를 인식하기 위해서 Tesseract-OCR[7]을 사용하였다. 동심 원당 작은 원의 개수 중 가장 많은 동심원을 선택하여 아날로그 게이지 숫자의 위치를 찾아 Tesseract-OCR을 적용한다. 마지막으로 인식된 숫자와 아날로그 게이지의 지침의 각도를 이용하여 현재의 아날로그 게이지의 값을 산출한다.

두 번째 방법은 Fully Convolutional Network(FCN)를 적용한 방법이다. 먼저 아날로그 게이지 이미지를 사용하여 FCN을 학습시킨다. 학습시킨 데이터는 ICDAR (2015)[10] 275장과 직접 제작한 아날로그 게이지 데이터 275장을 조합하여 총 550장의 데이터를 사용하였다. 학습한 FCN으로 물체 탐지(object detection)를 수행하며 아날로그 게이지의 숫자나 글자가 있을 가능성이 높은 영역을 추출한다. Tesseract-OCR에 추출된 영역을 적용시켜 아날로그 게이지의 숫자와 문자를 판독하게 된다.

1.2 논문 구성

본 논문은 5장으로 구성되어 있다.

2장에서는 두 알고리즘에 사용된 영상처리 방법과 인공 신경망에 대해 기술한다.

3장에서는 두 알고리즘을 설명한다.

4장에서는 하드웨어 구성을 설명하고 실험의 결과를 분석한다.

5장은 요약과 결론을 맺는다.

2. 연구의 이론적 배경

2.1. 디지털 영상처리

디지털 영상처리[4, 5]는 기계로 영상 정보를 생성, 처리, 인식등 다양한 기술을 적용하는 분야이다. 대표적으로 두 가지의 분야가 기반이며 인간이 영상 정보를 획득하기 편하게 가공하는 이미지 프로세싱(Image Processing)과 기계가 영상 정보를 인식할 수 있도록 영상 정보를 처리하는 컴퓨터 비전(Computer Vision)이 있다. 두 분야의 기술은 서로 판이하지 않고 목적에 따라 구분되는 것이므로 본 논문의 서술에서는 별도로 구분하지 않는다.

2.1.1 이진화 임계 분할(Binarization Threshold)

이미지는 2차원 행렬이며 원소마다 다양한 값으로 존재하며 픽셀로 저장된다. 영상 처리를 수행할 때 모든 픽셀의 값이 필요하지 않을 경우 불필요한 부분을 제거할 필요가 있다. 이것을 영상 분할이라고 한다. 영상 분할을 하는 방법 중 하나가 이진화 임계 분할이며 문턱값(threshold)을 이용하여 0과 1로 바꾸는 방법이다. 이진화 임계 분할은 식(2.1)으로 나타낼 수 있으며 그림2.1은 이진화 임계 분할의 결과이다.

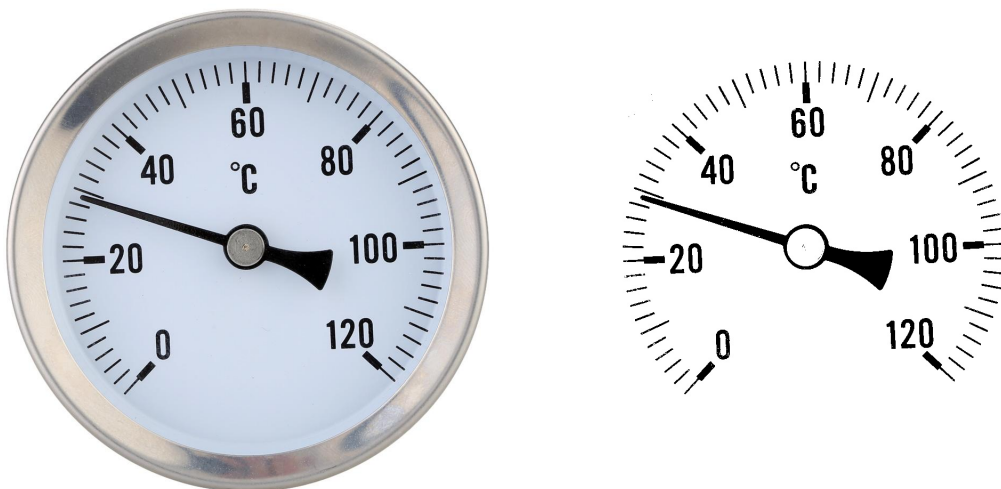


그림2.1 원본이미지(좌)와 이진화 임계 분할 결과(우)

$$g(x,y) = \begin{cases} 1 & f(x,y) > T \\ 0 & f(x,y) \leq T \end{cases} \quad (2.1)$$

이진화는 0과 1로만 이루어져 있으나 이미지의 표시를 위해 식(2.2)와 같이 1은 픽셀의 최대값인 255로 바꾸어 표기한다.

$$g(x,y) = \begin{cases} 255 & f(x,y) > T \\ 0 & f(x,y) \leq T \end{cases} \quad (2.2)$$

2.1.2 모폴로지(Morphology)

모폴로지는 영상 정보의 형태를 직접적으로 변형하는 디지털 영상처리의 기술이다. 모폴로지는 4가지의 방법으로 구성되어 있으며 침식(Erosion), 팽창(Dilation), 열림(Opening), 닫힘(Closing)이다. 4가지의 연산 방법은 논리연산 AND나 OR과 매우 흡사한 모습을 보인다. 일반적으로 모폴로지는 이진화된 이미지에서 수행한다.

침식은 마스크를 만족하는 픽셀이 모두 있을 경우만 표시하는 방법이다.. 침식의 효과로 물체가 줄어든다. 침식은 식(2.3)으로 표현하며 그림2.2 침식의 예시이다.

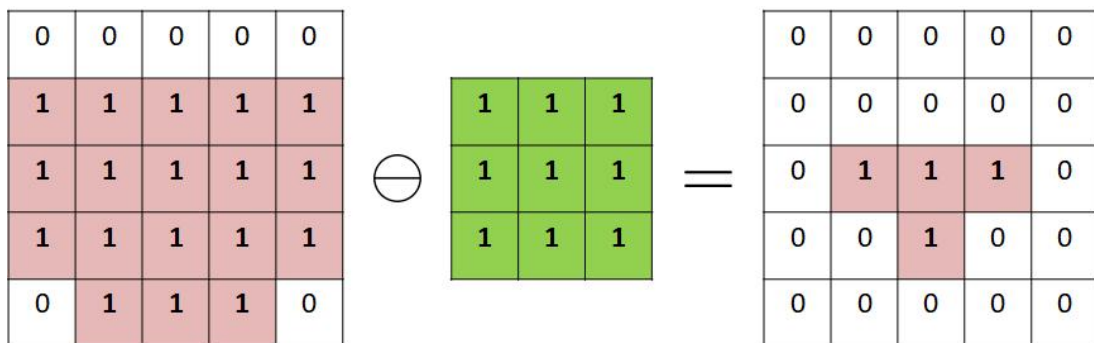


그림2.2 침식 연산의 예시

$$A \ominus B = \{z | (\hat{B})_z \cap A \neq \phi\} \quad (2.3)$$

팽창은 마스크를 만족하는 픽셀이 하나라도 있을 경우만 표시하는 방법이다.. 팽창의 효과로 물체가 늘어든다. 팽창은 식(2.4)으로 표현하며 그림2.3 침식의 예시이다. A는 이미지이며 B는 마스크이다.

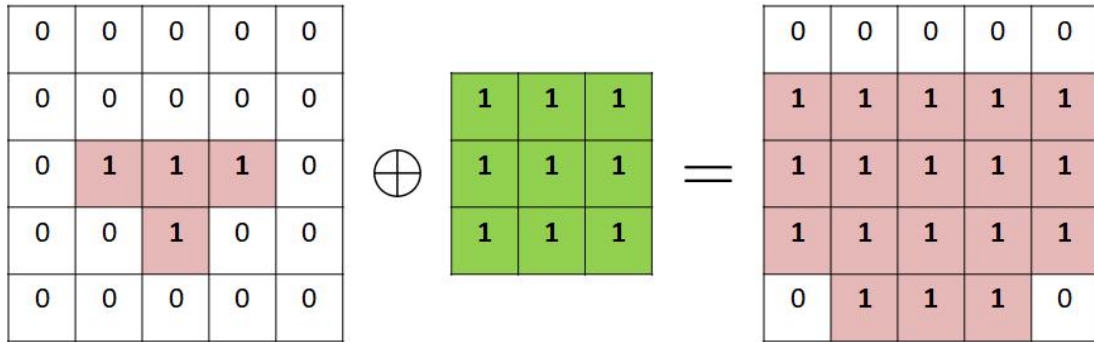


그림2.3 팽창 연산의 예시

$$A \oplus B = \{z | (B)_z \subseteq A\} \tag{2.4}$$

열림은 침식을 적용하고 그 결과에 팽창을 적용시키는 방법이다. 열림은 식(2.5)으로 표현하며 그림2.4 열림의 예시이다. 열림의 특징은 물체의 크기 변화가 없으며 마스크 보다 작은 영역만 삭제된다.

$$A \circ B = (A \ominus B) \oplus B \tag{2.5}$$

닫힘은 팽창을 적용하고 그 결과에 침식을 적용시키는 방법이다. 닫힘은 식(2.6)으로 표현하며 그림2.5 닫힘의 예시이다. 닫힘의 특징은 물체의 크기 변화가 없으며 마스크 보다 작은 영역은 매끈다.

$$A \cdot B = (A \oplus B) \ominus B \tag{2.6}$$

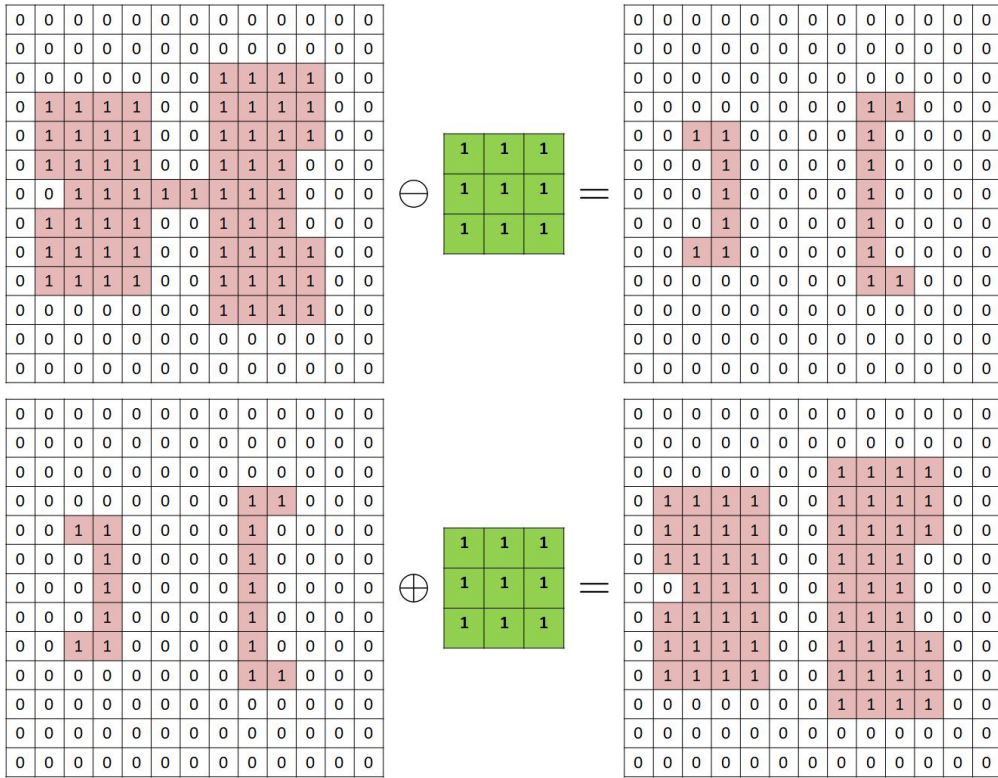


그림2.4 열립 연산의 예시

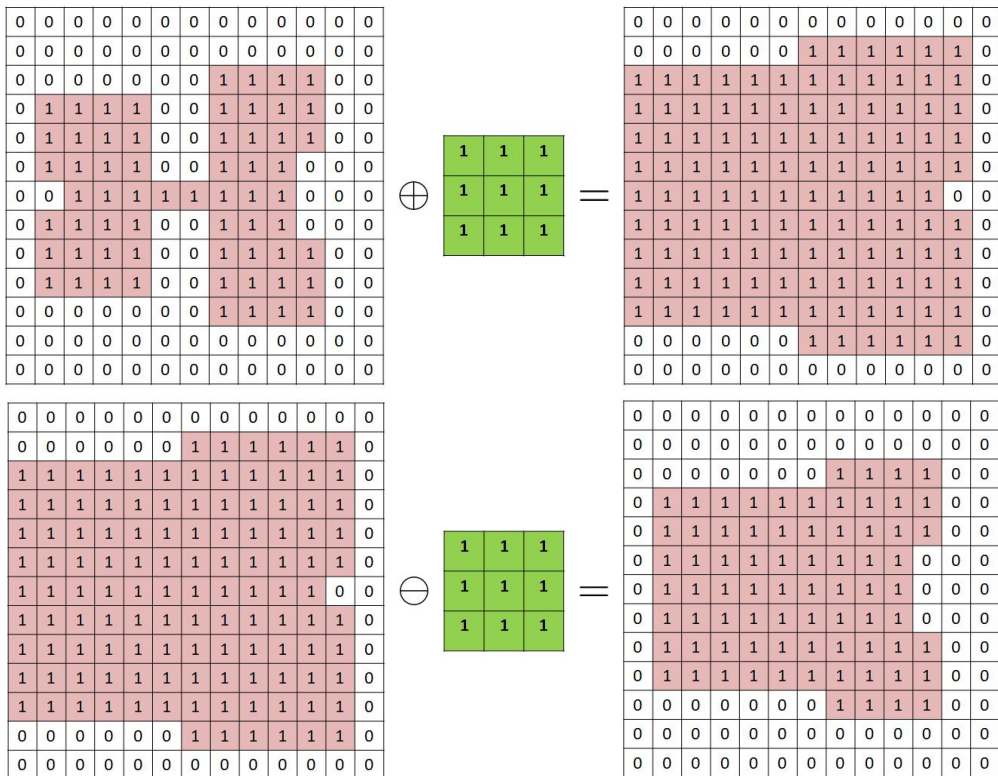


그림2.5 닫힘 연산의 예시

2.1.3 컨투어(Contour)

컨투어는 이미지 내에서 같은 값을 갖는 지점들을 연결한 것이다. 픽셀 값의 급격한 변화량인 에지(edge)를 찾는 데 사용할 수 있다.

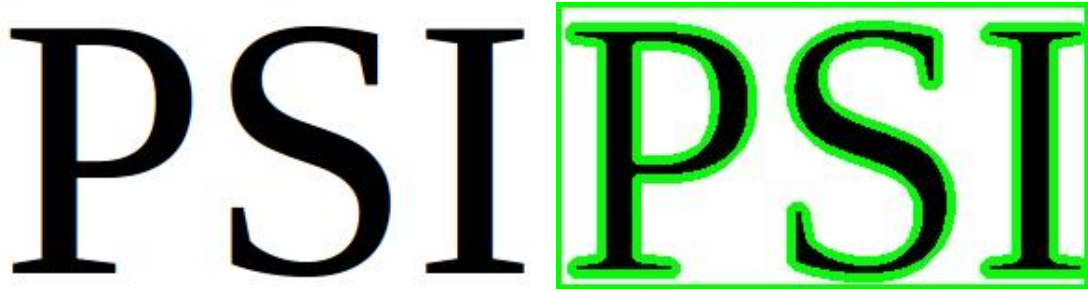


그림2.6 원본이미지(좌)와 컨투어 적용 결과(우)

2.1.4. 허프 변환(Hough Transform)

허프 변환[6]은 특징을 추출하는 기술로 영상처리, 디지털 이미지 프로세싱, 이미지 분석에서 사용한다. 허프 변환이 추출 가능한 특징은 대표적으로 직선이나 원이 있다. 일반 적으로 직선은 그림2.7처럼 구성되어 있으며 식(2.7)으로 표현한다. b 는 y 의 절편이며 a 는 기울기이다.

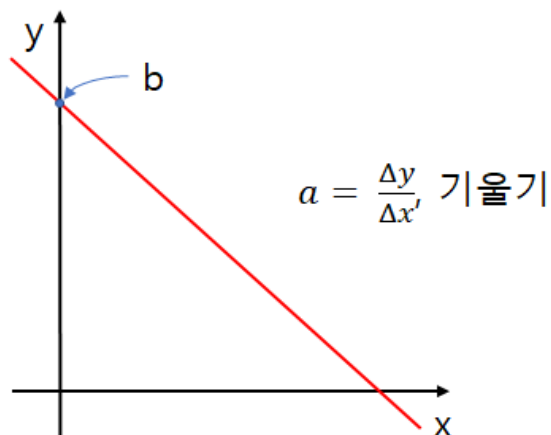


그림2.7. 직선의 구조

$$y = ax + b \quad (2.7)$$

이미지의 무수히 많은 점들 중 식(2.7)을 만족하면 직선의 구성 요소인 점이라 볼 수 있다. 그러나 무수히 많은 점들을 하나하나 확인하여 직선을 찾는 것은 매우 비효율적이다. 이러한 문제점을 해결하기 위해 허프 변환이 고안되었다. 식(2.7)에서 a 와 b 는 고정된 값이고 x 와 y 는 변화하는 값이다. 이것을 바꾸어서 x 와 y 를 고정 값이라고 하고 a 와 b 는 변화하는 값이라고 한다면 ab 평면의 한 점은 그 점이 가지는 모든 직선이다. 식(2.8)은 ab 평면으로 전환된 직선의 방정식을 나타낸다.

$$b = -xa + y \quad (2.8)$$

ab 평면의 직선은 한 점이 가지는 모든 직선을 의미한다. 그림 2.8처럼 만약 ab 평면에서 두 직선이 만난다면 이를 만족하는 직선은 오직 하나 이므로 그 점은 직선이 된다.

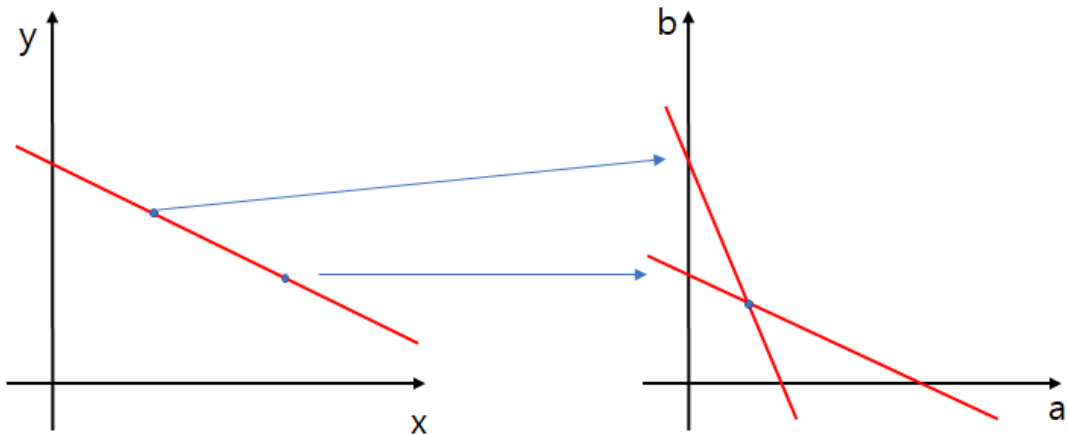


그림 2.8 xy 평면과 ab 평면의 관계

단순히 평면을 바꾸는 것으로 직선을 쉽게 찾을 수 있다. 다만 기울기 값은 무한대이므로 ab 평면에 표현할 수가 없다. 그래서 식(2.9)과 그림 2.9 같이 변환하여 사용한다.

$$r = x\cos\theta + y\sin\theta \quad (2.9)$$

$$r \in R, \theta \in [0, 180]$$

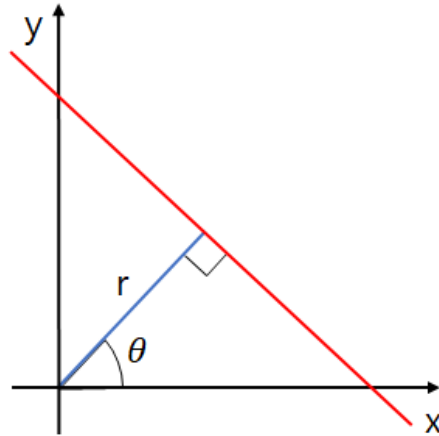


그림2.9 직선을 r과 θ 의 변환

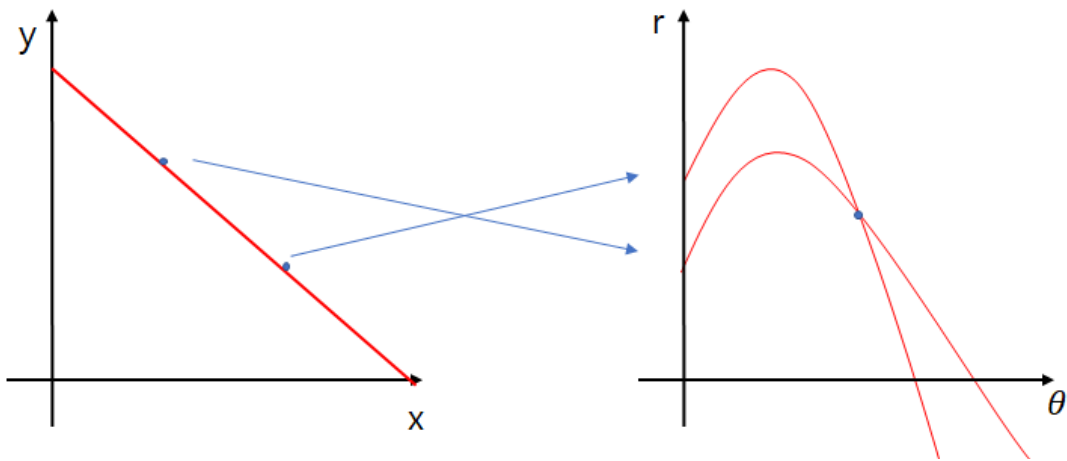


그림2.10. xy 평면과 $r\theta$ 평면의 관계

그림2.10은 허프 변환의 직선 검출 과정을 보여준다. $r\theta$ 평면에서 두 곡선이 만난다면 이를 만족하는 직선은 오직 하나 이므로 그 점은 직선이 된다.

일반적으로 원은 그림2.11과 식(2.10)으로 표현한다. a 와 b 는 원의 중심점이며 R 은 원의 반지름이다.

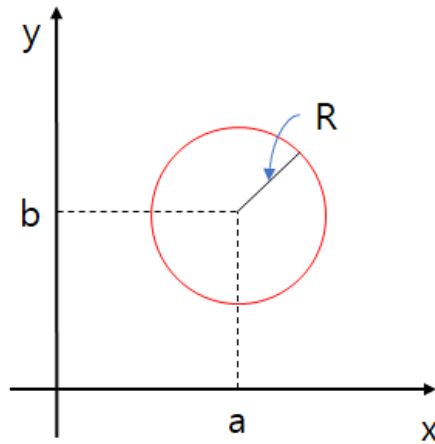


그림2.11 원의 구조

$$(x - a)^2 + (y - b)^2 = r^2 \quad (2.10)$$

식(2.10)은 x 와 y 에 대한 원의 방정식이다. 직선에 대한 허프 변환과 마찬가지로 무수히 많은 점들을 하나하나 확인하여 원을 찾는 것은 불가능하다. 원에 대한 허프 변환은 xy 평면에서 ab 평면으로 전환하여 임의의 점(x,y)에 대해 반지름이 R 인 원을 그린다. 모든 점에 대한 원이 가장 많이 겹치는 곳이 찾아야 하는 원의 중심이된다. 식(2.11)과 식(2.12), 그림2.12는 원에 대한 허프 변환을 나타낸다.

$$\begin{aligned} x &= a + R * \cos(\theta) \\ y &= b + R * \sin(\theta) \end{aligned} \quad (2.11)$$

$$\begin{aligned} a &= x - R * \cos(\theta) \\ b &= y - R * \sin(\theta) \end{aligned} \quad (2.12)$$

점이 가질 수 있는 모든 원을 표시하고 가장 많이 겹치는 곳이 중심점이다. 허프 변환으로 직선과 원을 보다 효율적으로 찾을 수 있다. 그러나 허프 변환은 수식을 기반으로 동작하므로 허프 변환이 검출을 할 수 있는 경우는 수식으로 표현이 가능

한 직선이나 원으로 한정된다. 또한 에지(edge)기반이므로 노이즈와 에지의 강도 등의 이유로 에지가 불확실하면 찾을 확률이 매우 낮아진다.

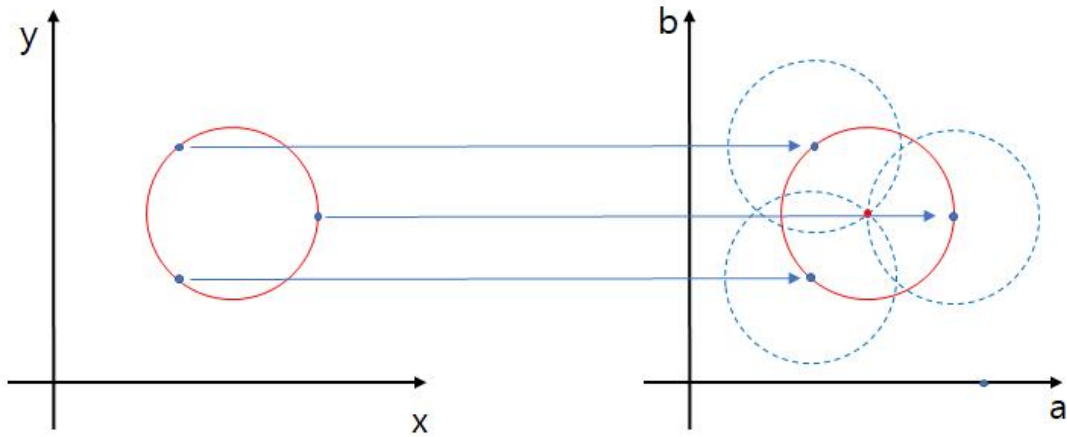


그림2.12 원에 대한 허프 변환

2.2 심층학습(Deep Learning)

심층학습은 기계학습의 방법 중 하나이며 인공 신경망을 기반으로 한다. 뇌의 구조에서 영감을 받은 심층학습은 정리되지 않거나 구조화 되지 않은 데이터로부터 자율학습을 할 수 있다. 기존 인공신경망의 모델의 단점을 극복과 하드웨어의 발전, 빅데이터로 인해 근래 급격하게 발전하고 있다. 심층학습은 자체적으로 데이터의 특징을 추출하여 학습에 사용하므로 데이터의 특징을 인간이 직접 선정하여 학습시키는 기계학습(Machine Learning)보다 진보한 개념이다. 최근 이미지 인식, 분류 등 다양한 분야에서 우수한 성능으로 많이 사용되고 있다. 대표적으로 컨볼루션 신경망(Convolution Neural Network)이 있다.

2.2.1. 이미지 인식(Image Recognition)

이미지 인식은 어떤 기계에게 이미지 상의 물체 혹은 대상이 무엇이고 어디에 있는지를 파악하도록 하는 것이다. 예를 들어 사람이 그림2.13을 본다면 사과가 이미지

안에 있다는 것을 파악할 수 있다. 그러나 기계의 입장에서는 무수히 많은 픽셀로 이루어진 행렬에 불가하다. 사람은 이미 학습된 사과라는 물체를 어떠한 매커니즘으로 사과의 정보를 꺼내어 판단을 한다. 이러한 사과 인식 과정은 매우 빠르게 진행된다. 반면 기계는 기술이 월등히 발달한 현재에도 이러한 수준에 도달하지 못하였다.



그림2.13 나무에 사과가 있는 이미지

기계의 물체 인식 성능을 사람의 수준으로 만들기 위해 많은 노력을 하였다. 뇌가 물체를 인식하는 방식을 모방하여 기계를 학습 시킨 방법이 뉴럴 네트워크(Neural Network)이다. 예를 들어 사과를 인식하기 위해서는 사과의 색상과 모양이 있을 것이다. 이를 적절히 특징을 뽑아내고 어떠한 알고리즘에 넣는다면 기계는 사과를 인식할 수 있을 것이다. 이미지 인식은 다양하고 방대한 데이터를 뉴럴 네트워크와 같은 우수한 알고리즘과 결합하여 사물을 판단하는 것이다.

이미지 인식에는 분류(Classification), 탐지(Detection), 시맨틱 세그멘테이션(Semantic segmentation)이 있다. 분류(그림2.14)는 입력 이미지의 단일 사물이 어떤 클래스에 해당되는지 판단하는 것이다. 다수의 학습된 클래스중 해당 사물이 어느 클래스인지 평가를 하고 그중 점수가 가장 높은 것으로 해당 클래스를 판단한다. 입력 이미지에 다수의 물체가 있다면 가장 확률이 높은 것을 해당 클래스라 판단한다.



그림2.14 심층 학습을 이용한 이미지 분류의 예시

탐지(그림2.15)는 입력 이미지 안의 어떤 사물이고 어디에 있는지를 판단하는 것이다. 분류에서 위치를 찾는 문제가 추가 되었다. 일반적으로 사물은 바운딩 박스(Bounding box)의 형태로 표시한다.

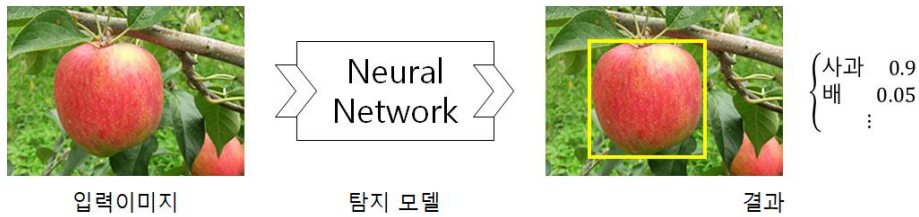


그림2.15 심층 학습을 이용한 이미지 탐지의 예시

시맨틱 세그멘테이션(그림2.16)은 입력 이미지의 사물이 학습된 클래스에 해당된다면 어디에 있는지 픽셀 단위로 표시하는 것이다. 물체의 위치를 박스 하나로 표시하며 대략적인 위치만을 파악하는 탐지에 비해 시맨틱 세그멘테이션은 물체의 정확한 위치를 파악해야 하므로 연산량이 많다.

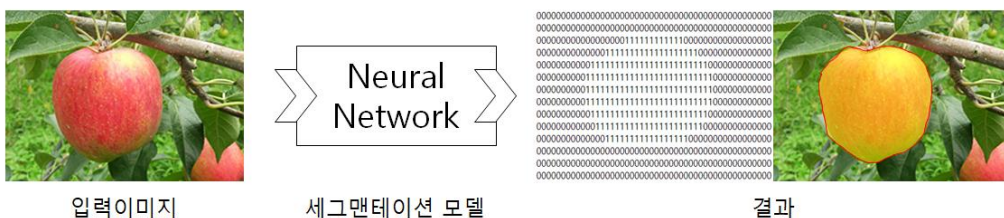


그림2.16 심층 학습을 이용한 이미지 시맨틱 세그멘테이션의 예시

2.2.2. 퍼셉트론(Perceptron)

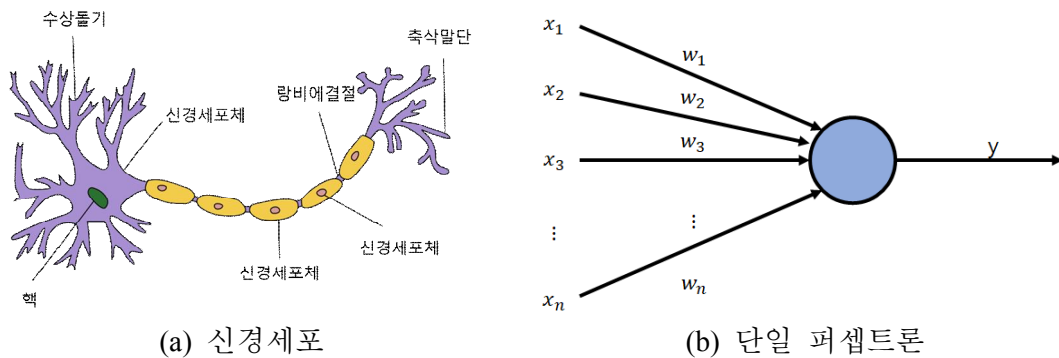


그림 2.17 신경세포와 단일 퍼셉트론

퍼셉트론[11]은 인공신경망(Artificial Neural Network)중 하나이며 딥러닝의 가장 기초 기술이다. 퍼셉트론은 1957년에 Frank Rosenblatt가 동물의 신경계를 본 따 만들었다. 그림2.5(a)에 나온 신경세포(Neuron)는 시냅스를 통해 다른 신경세포로 전기적 신호를 전송한다. 신호를 받은 신경세포는 신호의 세기에 따라 세포를 활성화한다. 그림2.5(b)의 퍼셉트론은 이러한 신경세포의 구조를 모방하여 입력마다 다른 가중치(Weight)를 받아서 값이 1이 넘을 경우 출력을 하게된다. 식(2.13)은 퍼셉트론을 표현했다. x 는 입력, y 는 출력, w 는 가중치이며 b 는 바이어스, $h(x)$ 는 활성화 함수이다.

$$y = \begin{cases} 0 & (b + w_1x_1 + w_2x_2 + \dots) \leq 0 \\ 1 & (b + w_1x_1 + w_2x_2 + \dots) > 0 \end{cases}$$

$$y = h(b + w_1x_1 + w_2x_2 + \dots) \quad (2.13)$$

$$h(x) = \begin{cases} 0 & (x \leq 0) \\ 1 & (x > 0) \end{cases}$$

퍼셉트론은 선형이므로 AND나 OR 연산과 같이 하나의 직선으로 나눌 수 있는 문제에 적합하지만 XOR과 같은 하나의 직선으로 나눌 수 없는 문제는 적용이 불가능하다.

하나의 직선으로 나눌 수 없는 문제를 해결하기 위해 다층 퍼셉트론(Multi-Layer Perceptron, MLP)이 제시되었다. 다층 퍼셉트론은 퍼셉트론을 여러 개 쌓아 각층을 연결한 모양이다. 다층으로 구성되었으므로 복잡한 문제를 해결할 수 있다.

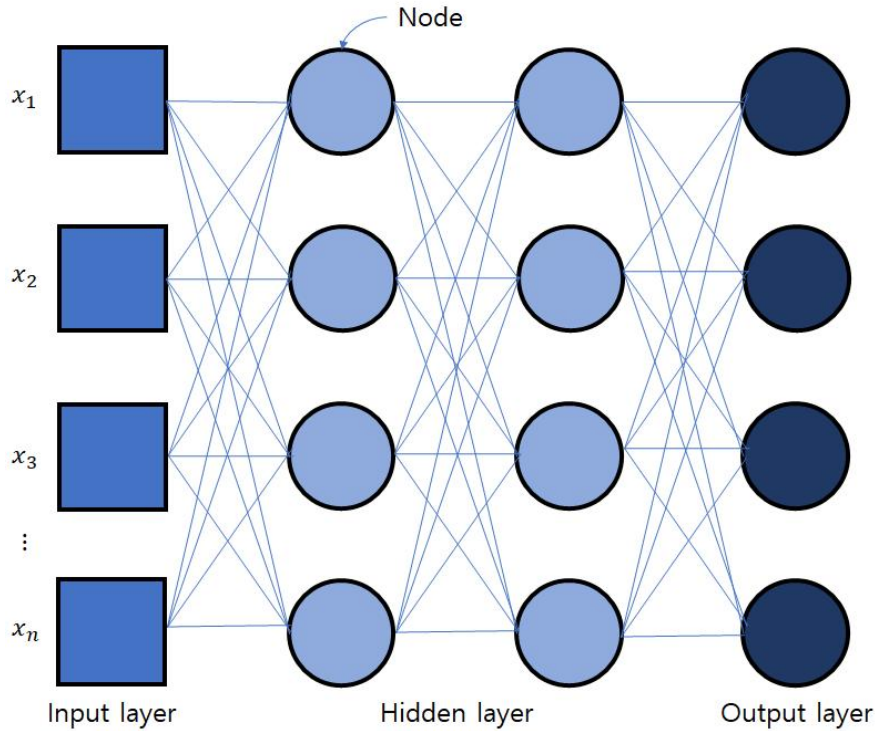


그림 2.18 다층 퍼셉트론

입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 이루어져 있다. 은닉층의 인공 뉴런(노드, node)은 가중치를 내포하고 있으며 입력이 인공 뉴런을 통과하면서 가중치가 곱해지고 합산되어 결과가 도출된다.

단일 퍼셉트론의 노드는 하나이므로 가중치를 사용자가 수동으로 결정하더라도 어려움이 없었다. 반면 다층 퍼셉트론은 노드가 증가하며 각 노드가 유기적으로 연결되어 있다. 이것은 층이 늘어날수록 사용자의 작업량이 늘어남을 의미한다. 많은 노드를 수동으로 조정하기 어려움으로 역전파등의 기술로 가중치를 스스로 결정한다. 다층 퍼셉트론은 인공신경망과 같은 의미이다.

2.2.3. 활성화 함수(Activation Function)

활성화 함수는 입력 신호의 합을 그대로 사용하지 않고 출력 신호를 변환하는 함수이다. 입력 신호의 합을 활성화할지 밀지를 결정하는 함수이다. 활성화 함수는 비선형 함수를 사용하는데 비선형 함수를 사용하지 않으면 은닉층을 늘려도 효과가 없다. 예를 들어 활성화 함수가 선형 함수 $h(x) = ax$ 이고 은닉층이 2단계라면 $y(x) = h(h(x))$ 이다. 이 식은 $y(x) = 2ax$ 가 된다.

- 계단 함수(Step Function)

특정 임계 값을 넘기면 활성화되는 함수이다. 쉽게 예측이 가능한 함수지만 실제로 잘 사용되지 않는데 출력 값이 0 아니면 1값으로 극단적이므로 은닉층에서 값의 손실이 크게 일어나게 된다. 그림 2.19와 식(2.14)는 계단 함수를 나타낸다.

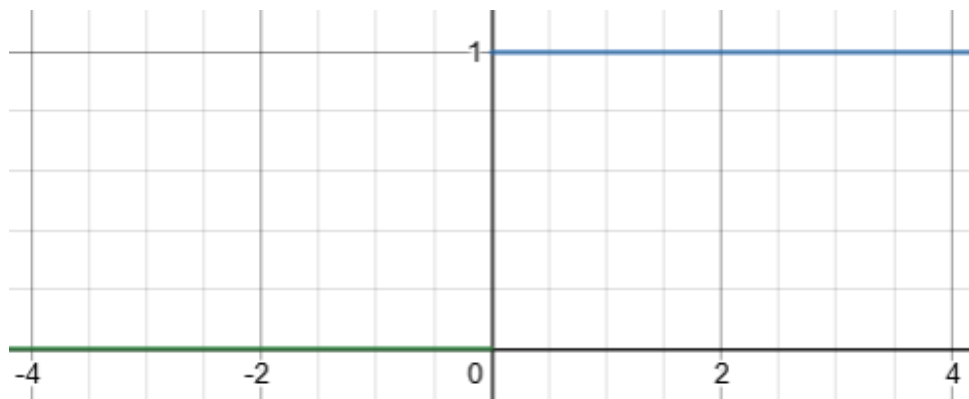


그림 2.19 계단 함수

$$f(x) = \begin{cases} 0 & (x \leq 0) \\ 1 & (x > 0) \end{cases} \quad (2.14)$$

- 시그모이드 함수(Sigmoid Function)

시그모이드 함수는 1과 0의 출력을 갖는다. 하지만 극단적으로 값이 바뀌는 계단 함수와는 달리 다소 완만하게 변하며 값을 잃지 않기에 계단 함수보다 우수하다. 그러나 역전파를 계산할 때 Vanishing Gradient Problem을 일으킨다. 시그모이드의 양 끝의 값은 거의 0에 수렴한다. 0에 수렴하는 수는 미분의 곱인 역전파 알고

리즘에서 0으로 수렴시켜버려 쓸모없게 만들어 버린다. 그림2.20와 식(2.15)는 계단 함수를 나타낸다.

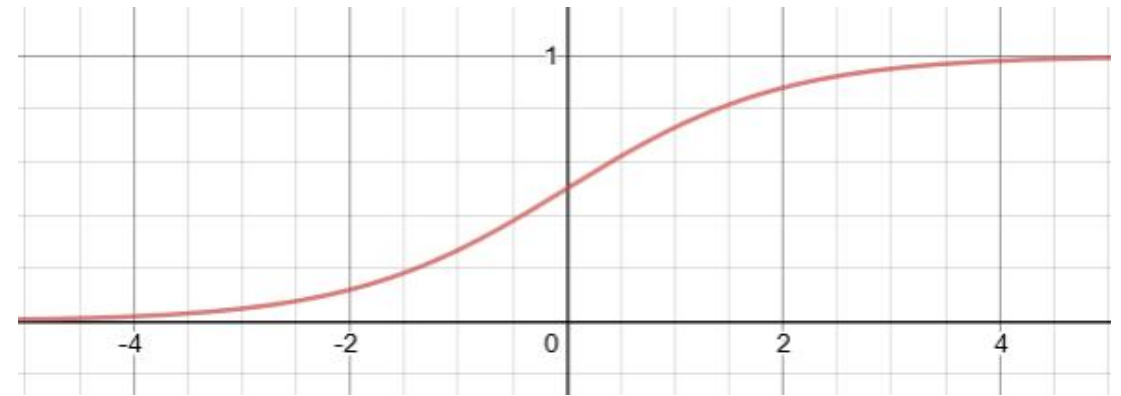


그림2.20 시그모이드 함수

$$f(x) = \frac{1}{1+e^{-x}} \quad (2.15)$$

- ReLU 함수(Rectified Linear Unit)

ReLU[12]는 입력이 0을 넘으면 입력을 유지하고 입력이 0을 넘지 못하면 0으로 출력하는 함수이다. ReLU는 미분 값이 오직 0과 1이므로 Vanishing Gradient Problem 이 크게 줄어든다. 그림2.21와 식(2.16)는 계단 함수를 나타낸다.

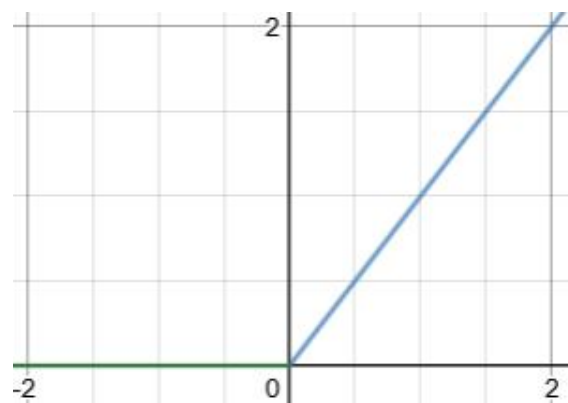


그림2.21 ReLU 함수

$$f(x) = \begin{cases} 0 & (x < 0) \\ x & (x \geq 0) \end{cases} \quad (2.16)$$

2.2.4 컨볼루션 신경망(Convolutional Neural Network, CNN)

컨볼루션 신경망은 컨볼루션 연산과 인공 신경망이 결합된 딥 러닝 기법이다. 시신경의 시각 처리 과정을 모방하여 설계 되었다. 인공 신경망은 1차원으로 이루어져 있다. 이러한 1차원 구조는 이미지와 같은 2차원 데이터에 적합하지 않았다. 예를 들어 이미지는 x, y의 좌표를 갖는 이차원 데이터 이므로 이를 처리하기 위해서는 1차원 인공신경망을 곱한형태로 처리해야한다. 이는 매우 많은 노드가 만들어지며 기계의 연산능력을 많이 요구하게된다. 이 문제를 해결하기위해 마스크라는 개념을 도입하였다. 마스크는 마스크를 만족하는 특징만 찾아낸다. 마스크를 사용함으로써 불필요한 데이터를 줄여 과적합을 방지하고 연산량도 줄일 수 있다. 그림2.22는 컨볼루션 신경망의 구조이다.

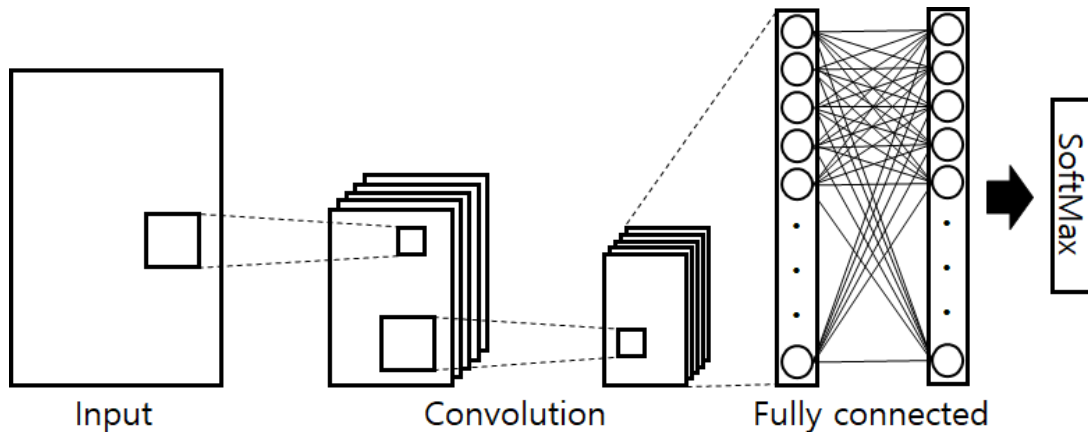


그림2.22 컨볼루션 신경망의 구조

컨볼루션 신경망은 컨볼루션 층(Convolution Layer), 풀링층(Pooling Layer), 전결합층(Fully Connected Layer), 소프트맥스(Softmax)로 구성되어있다. 컨볼루션은 이미지의 특징을 추출하며, 풀링층은 이미지의 특징을 골라내며, 전결합층은 판단을하고 소프트맥스는 결과를 확률로 나타낸다.

- 컨볼루션 층(convolution layer)

컨볼루션 연산은 커널에 대한 입력된 이미지의 반응이라 정의할 수 있다. 즉 커널의 모양에 따라 결과가 달라지며 특정 커널을 사용할 경우 이미지의 특징을 뽑아낼 수 있다. 커널과 입력 이미지를 1대1로 곱한 다음 모든 값을 합친다. 그림2.23과 식(2.17)은 컨볼루션 연산을 나타낸다. 식(2.17)의 n 은 픽셀 결과이며 w 는 커널의 값이다. 그림2.23의 커널은 y 축에 대한 변화를 알아보는 것이며 결과값 -8 은 중심 픽셀의 위쪽 픽셀이 아래쪽 픽셀보다 더 큼을 의미한다.

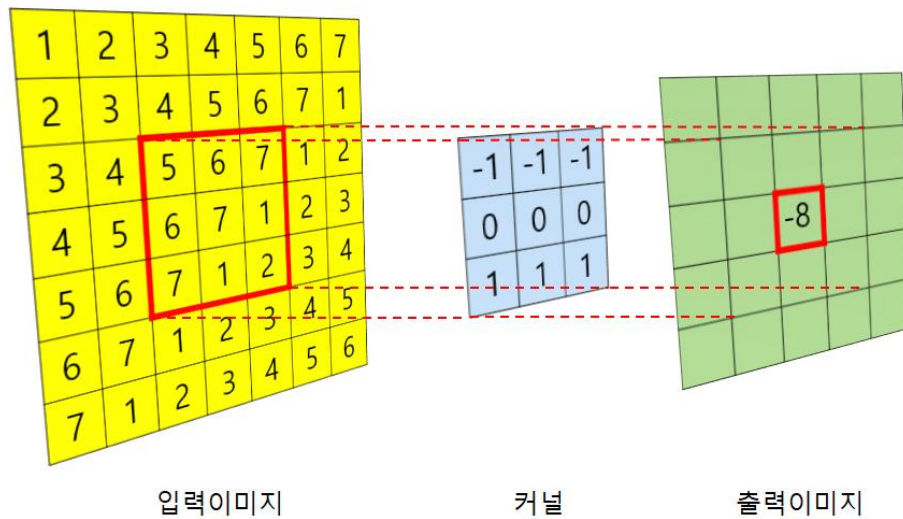


그림2.23 컨볼루션 연산

$$y(x,y) = \{w_1x_1 + w_2x_2 \dots w_nx_n\} \quad (2.17)$$

- 풀링층(pooling layer)

풀링층은 컨볼루션층을 통과하여 추출된 특징들 중 일부분만 선택하는 곳이다. 풀링은 데이터의 크기를 줄일 수 있기 때문에 처리에 필요한 자원을 줄일 수 있으며 데이터량이 너무 많아 모델이 데이터에 너무 맞는 과적합(Overfitting)을 줄일 수 있다. 풀링의 방법중에는 최대풀링(Max Pooling), 평균풀링(Mean Pooling)등이 있다. 컨

볼루션 신경망에서는 주로 최대풀링이 많이 사용되므로 이에 대한 설명을 한다. 최대 풀링은 영역 내에 최대 값을 뽑아낸다.

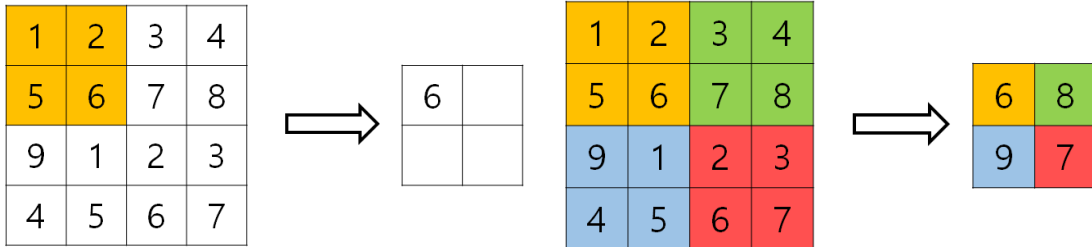


그림2.24 최대풀링 과정(좌), 최대풀링 결과(우)

그림2.24에서 필터의 크기는 2x2이며 필터의 이동간격(Stride)는 2이다. 최대풀링으로 4x4 크기의 입력 이미지가 2x2 크기로 줄어들었으며 각 필터에서 가장 큰 값만 추출한 것을 볼 수 있다. 풀링층 이후에는 앞서 설명한 활성화 함수에 의해 활성화 여부를 판단한다.

- 전결합층(Fully Connected Layer)

전결합층은 앞서 설명한 인공 신경망과 동일한 것이다. 다만 입력에 2차원 이미지므로 1차원 데이터로 바꾸기 위해 이미지를 한 줄씩 연결한다. 이것은 평탄화과정(flatten)이라고 한다.

- 소프트맥스(Softmax)

모든 층을 통과한 값을 분석하기 힘들다. 이러한 문제를 해결하기 위해 소프트맥스를 사용한다. 소프트맥스는 입력값을 0~1사이의 값으로 정규화하여 확률로 변환시킨다. 식은 소프트맥스를 나타낸다. 식(2.18)은 소프트맥스를 나타낸다.

$$y_k = \frac{e^{(a_k)}}{\sum_{i=1}^n e^{(a_i)}} \quad (2.18)$$

2.2.5. FCN(fully Convolutional Network)

컨볼루션 신경망의 단점은 입력 이미지의 고정과 위치 정보의 소실이다. 이러한 단점은 컨볼루션 신경망의 구조에 의해 발생한다. 컨볼루션층과 풀링층을 통과한 데이터가 전결합층으로 들어갈 1차원 데이터로 변환하는데 이때 2차원 데이터가 1차원으로 변환되고 전결합층에서 변환된다. 또한 전결합층은 데이터의 크기가 항상 같아야 하므로 2차원인 이미지 역시 크기가 같아야 한다. 이러한 단점으로 인해 컨볼루션 신경망을 이용하여 물체를 찾는 일에 적합하지 않다.

이러한 컨볼루션 신경망의 단점을 해결하기 위해 FCN이 고안되었다. 그림2.25와 같이 FCN은 전결합층 대신 1x1 컨볼루션 층이 배치되어있다. 즉, 모든 네트워크가 컨볼루션 층이다.

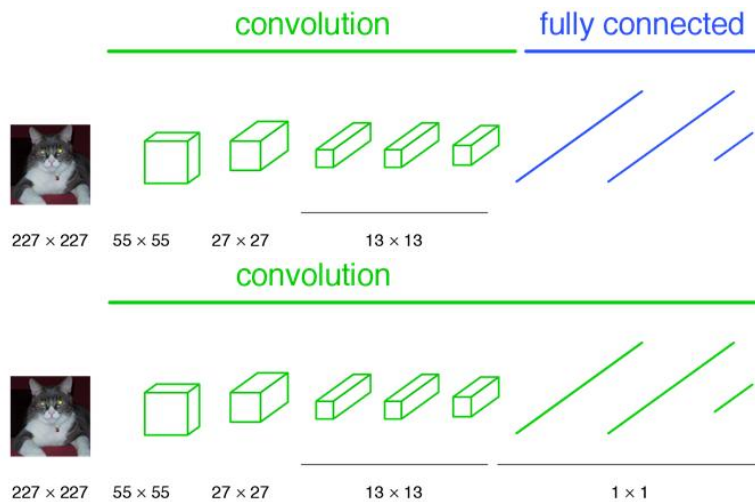


그림2.25 CNN과 FCN의 차이[2]

전결합층을 1x1 컨볼루션으로 바꿈으로써 위치 정보를 유지할 수 있으며, 모든 층이 컨볼루션 층이므로 입력 영상에 제한을 받지 않는다. 부가적으로 이미지를 한꺼번에 처리함으로써 처리 속도가 향상된다.

컨볼루션층과 풀링층을 거치면 마지막 컨볼루션 층은 입력 이미지에 비해 매우 작아진다. 시맨틱 세그멘테이션을 수행하기 위해 마지막 컨볼루션층의 크기를 다시 복원하는 과정이 필요하다. 주로 업샘플링(Upsampling)은 이중선형 보간법(Bilinear

Interpolation)을 사용한다. 하지만 이중선형 보간법은 학습된 값과 무관한 값을 넣는 것이므로 학습된 데이터가 아니다. 그림2.26은 FCN의 구조를 나타낸다.

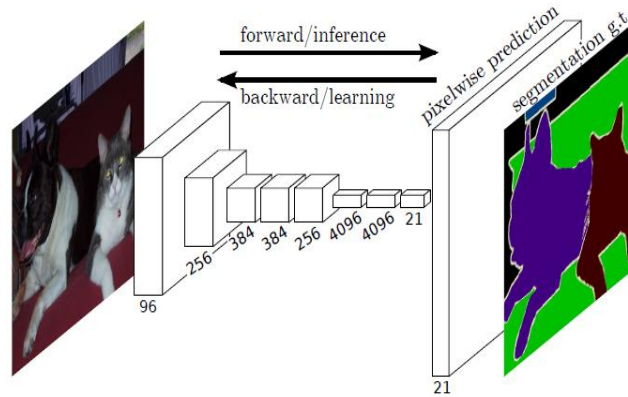


그림2.26 FCN의 구조[2]

FCN은 이미 학습된 이전 계층의 데이터를 업샘플링에 이용한다. 그림2.27과 같이 먼저 Pool5의 데이터를 2배 업샘플링한 결과와 Pool4의 데이터와 결합한다. 그 다음 합친 데이터를 2배 업샘플링하여 Pool3의 데이터와 결합한다. 3개의 컨볼루션층을 사용함으로써 보다 더 정확한 작업을 수행할 수 있다.

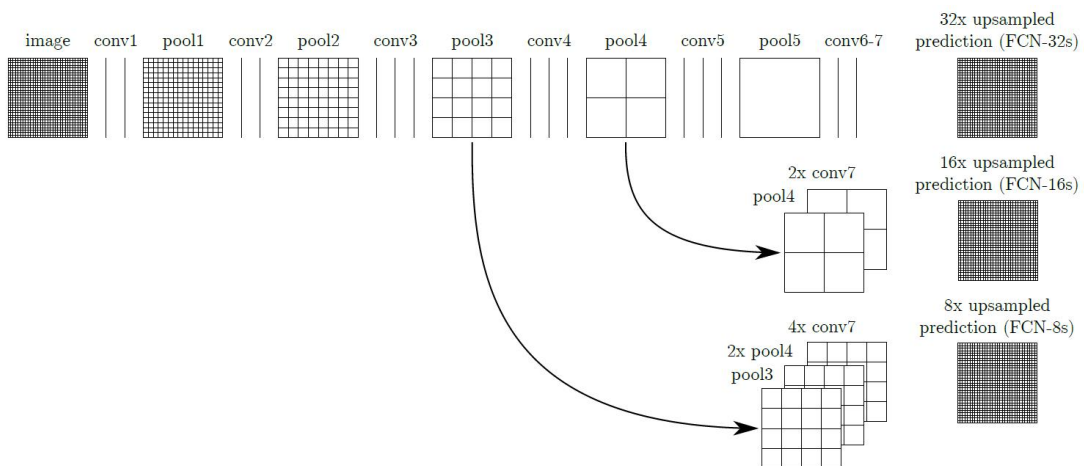


그림2.27 FCN의 업샘플링 방법[2]

2.2.6. 최소자승법

최소자승법은 어떤 데이터에 대해 규칙을 찾는 방법으로 데이터와 선택한 임의의 모델 값들 간의 차이를 제공하여 그 차이가 최소인 모델을 찾는 방식이다. 식2.29와 그림2.28은 최소자승법의 방법을 보여준다.

$$S = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (2.19)$$

y_i 는 데이터의 값을 나타내며 $f(x_i)$ 는 임의의 모델 값을 나타낸다. 모델 $f(x)$ 의 형태에 따라 직선이나 원등으로 데이터의 규칙을 찾을 수 있다.

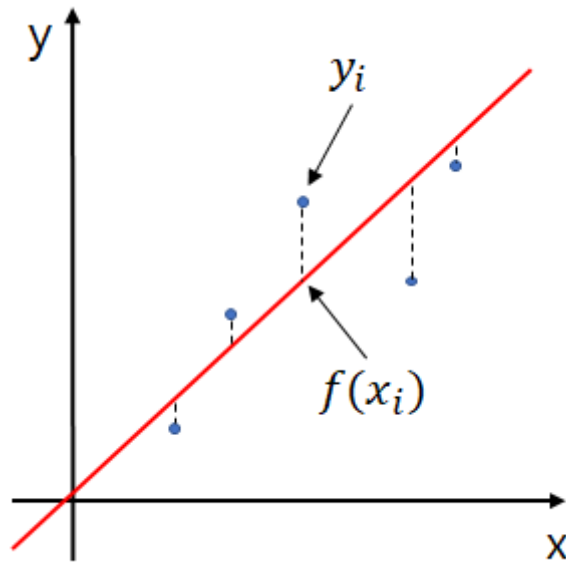


그림2.28 최소자승법의 원리

2.3. Tesseract-OCR

Tesseract-OCR[7]은 OCR라이브러리중 하나이며 1985년 HP에서 개발을 시작하고 2006년부터 구글에서 개발 및 배포한 오픈라이브러리이다. OCR(Optical Character Recognition)은 신문이나 책 같은 인쇄된 매체를 기계로 읽는 기술로 패턴인식, 인공지능, 컴퓨터 비전과 밀접한 관련있다. 그림2.28은 Tesseract-OCR의 예시이며 글자가 있는 이미지를 인식하여 출력한 것이다.

Brief history

Tesseract was originally developed at Hewlett-Packard Laboratories Bristol and at Hewlett-Packard Co, Greeley Colorado between 1985 and 1994, with some more changes made in 1996 to port to Windows, and some C++izing in 1998. In 2005 Tesseract was open sourced by HP. Since 2006 it is developed by Google.



Brief history

Tessemcl was originally developed al HewleIIrPackald Labomlories Brislol and al HewleIIrPa ckald Co, Greeley Colomdo between 1985 and 1994, with some more changes made in 199610 porl Io Windows, and some CHi Zing in 1998. In 2005 Tessemcl was open sourced by HP. Since 2006 H is developed by Google.

그림2.28 Tesseract-OCR의 적용 예시

3. 아날로그 게이지 정보 획득

본 연구에서는 아날로그 게이지의 정보를 인식하고 이를 판별하고자 한다. 제시하는 방법은 컴퓨터 비전과 Tesseract-OCR을 이용한 방법과 FCN과 Tesseract-OCR을 이용하는 방법이 있다.

3.1 영상처리를 이용한 아날로그 게이지 정보 획득

이번 장에서는 영상처리의 여러 알고리즘을 이용하여 아날로그 게이지의 정보를 획득하는 알고리즘[8]을 제시한다. 그림3.1은 시스템의 흐름도이다.

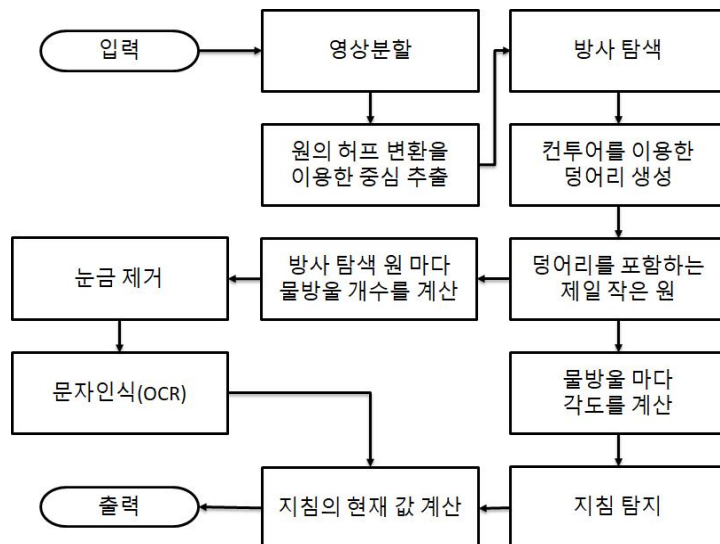


그림3.1 영상처리를 이용한 아날로그 게이지 정보 획득

아날로그 게이지는 지침이 동작하는 중심으로 부터 숫자가 일정한 간격으로 배치되어 있으며 지침은 반드시 중심을 지나가는 특징을 가지고 있다. 본 알고리즘은 규칙을 활용하여 아날로그 게이지의 중심부터 탐색하여 아날로그 게이지의 정보를 획득한다.

● 전처리

아날로그 게이지 이미지를 사용하면 노이즈로 인해 결과가 좋지 못하는 경우가 많다. 또한 아날로그 게이지는 사람이 사용하기 편하게 설계되어 가독성이 높기 때문

에 아날로그 게이지의 정보(숫자와 문자, 눈금, 지침)는 배경과 차이가 많이 난다. 이러한 특징을 활용하여 아날로그 게이지의 정보와 배경을 확실히 구분하기 위해 이진화 임계 분할을 사용하였다. 이진화 임계 분할만으로는 깔끔하게 분리할 수 없으므로 모폴로지 방법을 사용하였다.

모폴로지의 마스크는 3x3크기에 모양은 십자가 모양이며 열림을 적용하였다. 그림 3.2는 이진화 임계 분할의 결과를 보여준다.

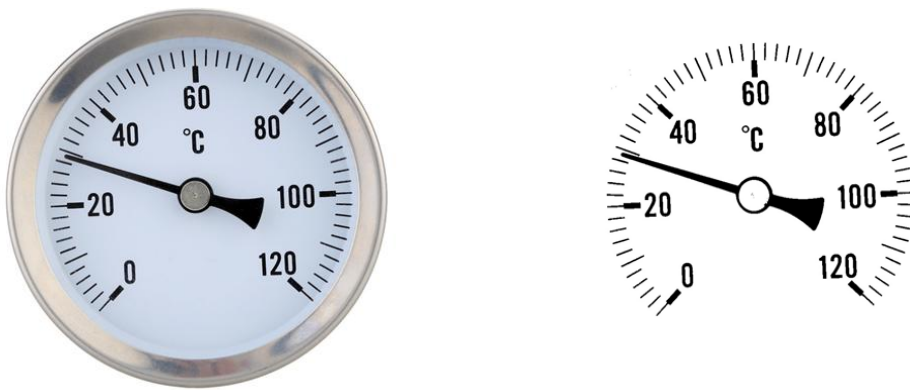


그림3.2 아날로그 게이지 이미지(좌)와 전처리의 결과(우)

- 원에 대한 허프 변환을 이용한 중심 추출

아날로그 게이지의 중심은 지침을 고정시키기 위한 원형태의 물체가 존재한다. 이를 탐지하기 위해 원에 대한 허프 변환을 사용하였다. 그림3.3은 원에 대한 허프 변환의 결과이다.

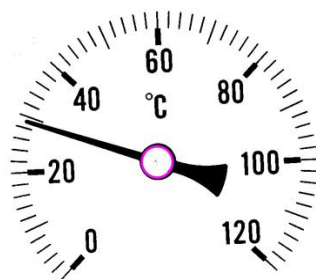


그림3.3 원에 대한 허프 변환의 결과

- 방사탐색

아날로그 게이지의 숫자와 눈금은 아날로그 게이지의 중심으로부터 원형으로 배치되어 있다. 이러한 특징들을 이용하여 아날로그 게이지의 중심으로부터 동심원을 그려 탐색을 한다. 그림3.4와 같이 동심원과 아날로그 게이지의 전처리된 아날로그 게이지 이미지가 만나는 지점을 추출한다.

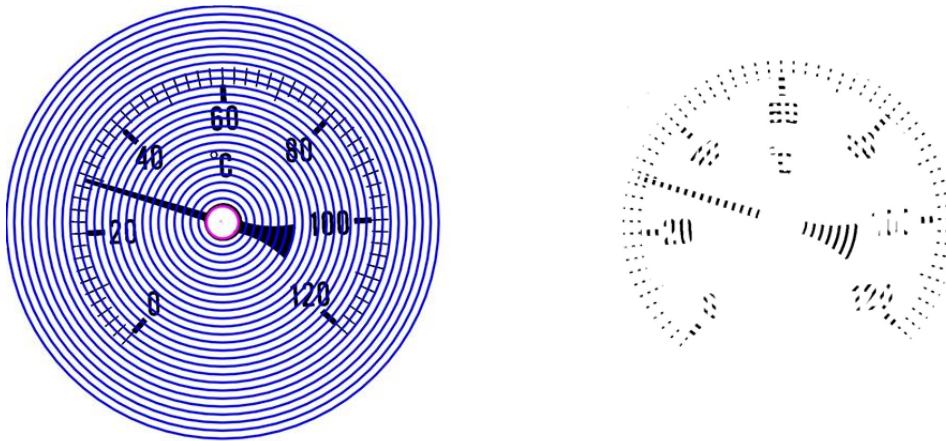


그림3.4 방사 탐색(좌)과 방사 탐색 결과(우)

- 아날로그 게이지의 지침 찾기

추출된 방사 탐색 결과에 대해 컨투어 방법을 적용하여 덩어리를 만든다. 덩어리의 모양이 불규칙하므로 그림3.5와 같이 덩어리를 포함하는 가장 작은 원을 만든다.

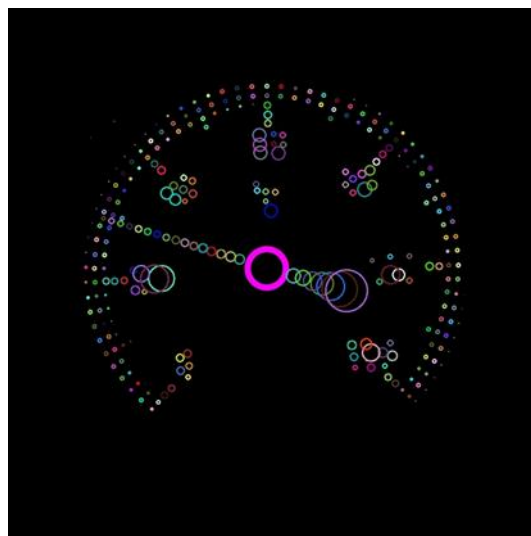


그림3.5 덩어리를 포함하는 가장 작은 원

아날로그 게이지의 지침은 길이가 가장 길며 아날로그 게이지의 중심을 지나는 직선이다. 이러한 특징을 이용하여 중심에서 방사탐색을 한다면 같은 각도로 검출이 될 것이다. 그림3.6는 각도에 따른 덩어리의 수 히스토그램을 나타내며 개수가 가장 많은 것 항목이 지침이다.

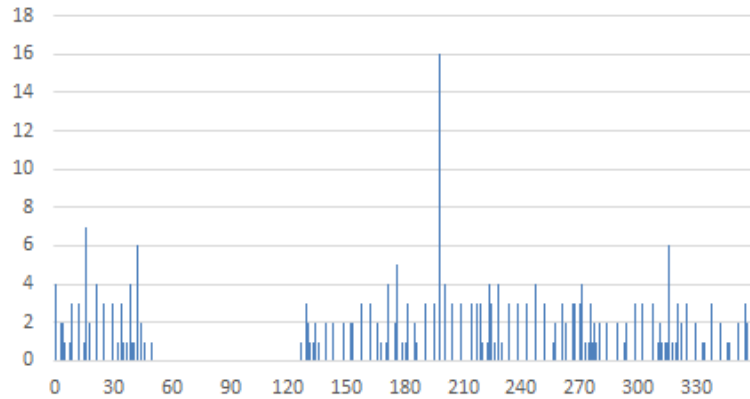
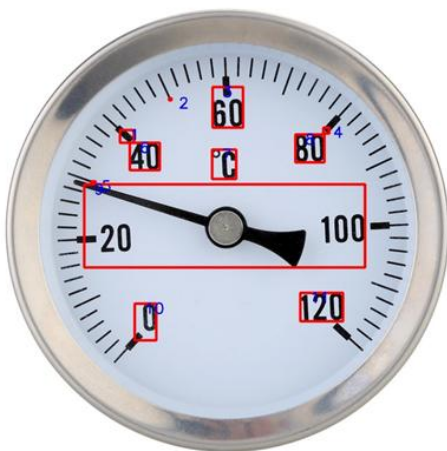


그림3.6 각도에 따른 덩어리의 수 히스토그램

● 아날로그 게이지 숫자의 관심 영역

아날로그 게이지의 숫자를 판독하기 위해 Tesseract-OCR을 적용하였다. 그러나 Tesseract-OCR은 신문이나 책과 같은 정렬이 잘된 문자만 인식하므로 그림3.7과 같이 아날로그 게이지의 숫자를 찾지 못하였다. 그 이유는 눈금과 숫자가 같은 검은 색이고 특징이 유사하기 때문이다.



- Box number : 1. word: '??'
- Box number : 2. word: '??'
- Box number : 3. word: '6'0'
- Box number : 4. word: '\'
- Box number : 5. word: '~'
- Box number : 6. word: '40'
- Box number : 7. word: '째[;'
- Box number : 8. word: '80';
- Box number : 9. word: 'W100'
- Box number : 10. word: '.0'
- Box number : 11. word: '120'

그림3.7 OCR 결과(좌)과 획득한 정보(우)

이 문제를 해결하고자 아날로그 게이지 숫자의 관심 영역을 설정하였다. 그림3.8와 같이 동심원 당 덩어리의 개수를 세어 가장 많은 곳으로부터 로컬(Local) 최저점을 찾는다. 개수가 가장 많은 지점은 눈금의 영역이다. 그 이유는 눈금의 개수가 게이지에서 개수가 가장 많고 눈금 다음에 존재하는 글자는 눈금보다 개수가 월등히 적기 때문에 로컬 최저점은 숫자와 눈금의 경계지점이다. 그림3.9과 같이 로컬 최저점에 해당하는 동심원기준으로 바깥쪽을 제거한다.

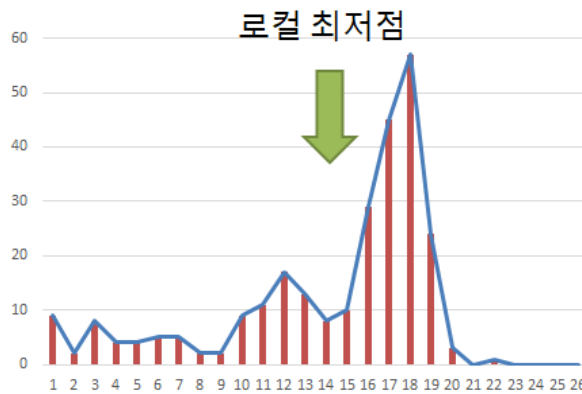


그림3.8 동심원에 따른 덩어리의 수 히스토그램

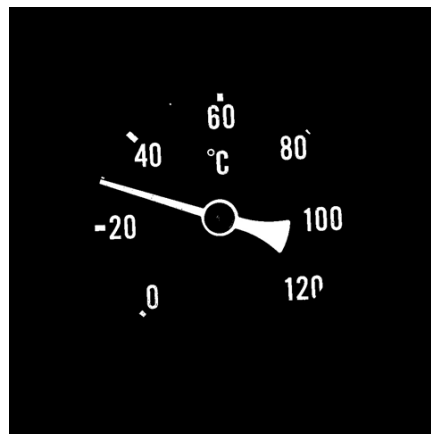
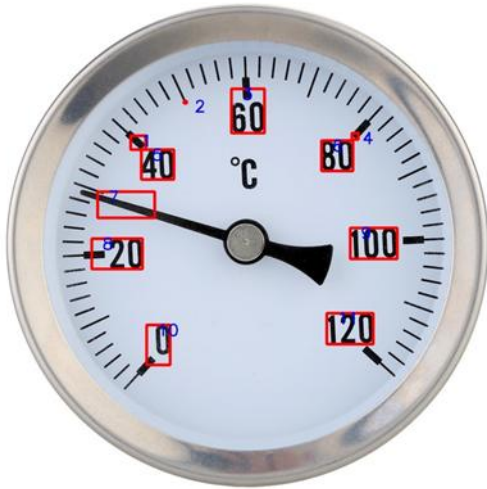


그림3.9 관심 영역 설정 결과



- Box number : 1. word: '??'
- Box number : 2. word: '??'
- Box number : 3. word: '6'0'
- Box number : 4. word: 'x'
- Box number : 5. word: '40'
- Box number : 6. word: '80'
- Box number : 7. word: '/'
- Box number : 8. word: '.20';
- Box number : 9. word: '100'
- Box number : 10. word: '.0'
- Box number : 11. word: '120'

그림3.10 OCR 결과(좌)과 획득한 정보(우)

그림 3.10과 같이 설정된 관심 영역에 Tesseract-OCR을 적용하여 숫자를 찾았다. 40과 80, 100, 120,을 찾을 수 있다.

● 각도 계산

그림3.11와 같이 숫자를 2개이상 찾는다면 각도 당 아날로그 게이지 값의 증가율을 계산할 수 있다. 식(3.1)은 아날로그 게이지 값의 계산식을 나타낸다. x 는 Tesseract-OCR로 찾아낸 숫자값이며 θ 는 숫자의 각도값이다.

$$ratio = \frac{x_1 - x_2}{\theta_1 - \theta_2} \begin{cases} x : \text{숫자 값} \\ \theta : \text{각도 값} \end{cases}$$

$$Value = x_1 - ratio(\theta_1 - \theta_0) \tag{3.1}$$

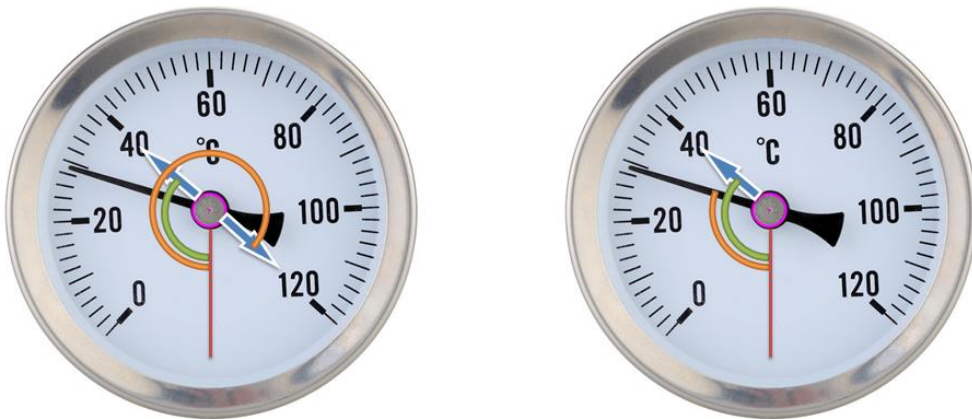


그림3.11 비율 계산(좌)과 각도 계산식(우)

3.2 FCN을 이용한 아날로그 페이지 정보 획득[9]

영상처리를 이용한 아날로그 페이지 정보 획득 방법은 사용되는 모든 기술이 사람이 직접 설정한 변수에 의해 작동되므로 다양한 모양의 아날로그 페이지에 대응이 어렵다. 이 문제를 해결하기 위해 FCN(Fully Convolutional Network)를 사용하였다. 아날로그 페이지에서 모호한 특징인 중심원을 찾는 것이 아니라 강력한 특징인 숫자부터 찾는 알고리즘으로 구성하였다. 그림3.9는 제시된 알고리즘을 보여준다. 아날로그 페이지 이미지를 훈련된 FCN에 넣어서 글자영역을 추출하고 Tesseract-OCR을 통해 판독한다. 최소자승법을 이용해 아날로그 페이지의 중심을 찾는다.

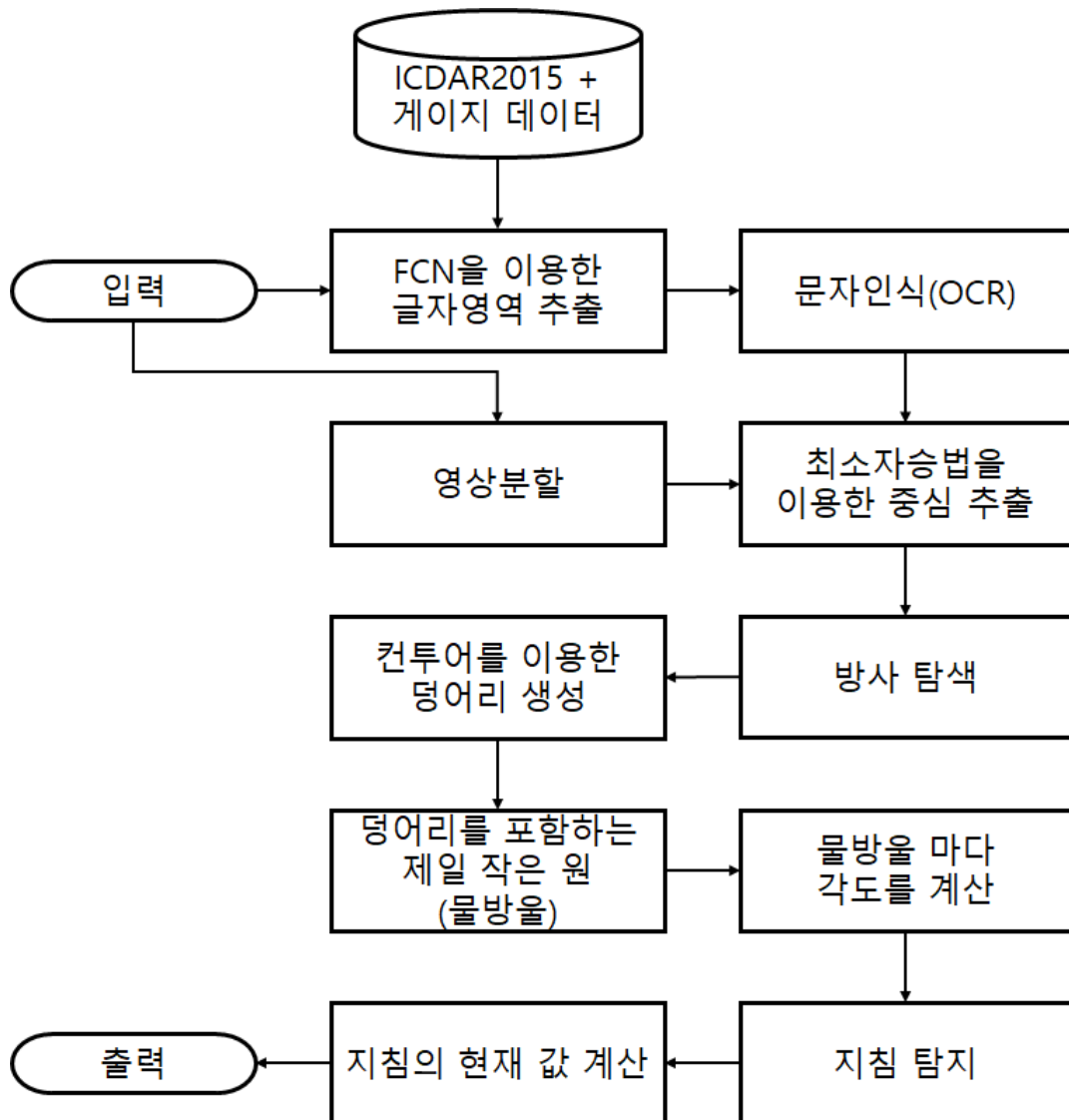


그림3.12 FCN을 이용한 아날로그 페이지 정보 획득

기존의 영상처리를 이용한 아날로그 페이지 정보 획득의 단점은 다음과 같다. 첫째 Tesseract-OCR이 아날로그 페이지의 글자를 인식하는데 어려움이 있다. 두 번째 원에 대한 허프변환으로 아날로그 페이지의 중심을 찾기 힘들다. 첫 번째 단점을 보완하기 위해 Fully Convolutional Network(FCN)을 이용하였다. FCN을 이용하여 글자의 영역을 추출한 다음 Tesseract-OCR이 판독을 한다. 두 번째 단점을 보완하기 위해 최소자승법을 사용한다. FCN과 Tesseract-OCR이 판독한 글자를 기준으로 원을 추정하여 아날로그 페이지의 중심을 찾는다. 비록 해당 방법은 아날로그 페이지의 정확한 중앙이 아니지만 원에 대한 허프트랜스폼보다 중심을 못 찾는 경우가 매우 줄어든다. 영상분할과 지침의 값을 계산하는 방법은 영상처리를 이용한 아날로그 페이지 정보 획득과 동일하므로 설명을 생략한다.

- FCN 훈련

FCN은 시맨틱 세그멘테이션에 특화된 네트워크이다. 눈금과 숫자가 가까이 있기 때문에 시맨틱 세그멘테이션을 수행하여 숫자를 정확히 탐지하고자 하였다. 시맨틱 세그멘테이션 후 바운딩 박스를 만드는 것까지 훈련하였다. 훈련에 사용된 데이터는 총 550장이며 ICDAR2015(International Conference on Document Analysis and Recognition)[10] 데이터 중 275장과 직접 라벨링한 아날로그 페이지 데이터 275장을 사용하였다. 두 데이터셋은 같은 구조로 이루어져 있으며 이미지 하나와 이미지 내부에 있는 숫자나 문자의 위치를 기록한 4개의 점과 해당 숫자나 문자의 값으로 이루어져 있다. ICDAR 데이터는 실생활에서 볼 수 있는 사진에 라벨링 된 데이터이다. ICDAR 데이터를 사용한 이유는 과적합 방지와 성능향상을 위해 사용하였다. 그림3.13은 ICDAR2015 데이터의 구성을 나타내고 그림3.14 아날로그 페이지 데이터의 구성을 나타낸다.



602,173,635,175,634,197,602,196,EXIT
734,310,792,320,792,364,738,361,I2R

그림3.13 ICDAR 데이터의 구성



78,171,93,169,93,185,78,187,20
82,121,98,119,98,136,82,138,40
111,78,126,76,126,92,111,94,60
154,56,169,54,169,69,154,71,80
194,66,216,64,216,79,194,82,100
218,102,240,98,240,113,218,116,120
223,149,244,146,244,161,223,164,140
204,194,225,190,225,206,204,209,160
206,215,218,212,218,220,206,223,###
139,209,177,204,177,210,138,216,###
139,216,182,210,182,217,138,223,###
138,225,159,221,159,227,138,230,###
139,245,190,237,190,246,139,254,###
103,212,110,210,110,224,103,226,0

그림3.14 아날로그 게이지 데이터의 구성

- Tesseract-OCR을 이용한 판독

FCN은 이용하여 추출된 숫자나 문자일 가능성이 높은 영역을 하나 씩 Tesseract-OCR을 적용하여 판독한다. 그림3.15는 Tesseract-OCR의 판독 결과를 나타낸다.



Box numer :1	Text :W
Box numer :2	Text :lirecau
Box numer :3	Text :Psi
Box numer :4	Text :40
Box numer :5	Text :60
Box numer :6	Text :30
Box numer :7	Text :10
Box numer :8	Text :20
Box numer :9	Text :501

그림3.15 Tesseract-OCR의 판독 결과

- 최소자승법을 이용한 중심 추출

최소자승법과 FCN이 찾아낸 글자들의 중심을 이용하여 아날로그 게이지의 중심을 찾아낸다. 글자의 중심이 기준이므로 아날로그 게이지의 중심과 정확히 일치하지 않는다. 때문에 오차가 발생하지만 기존의 원의 대한 허프트랜스폼은 아날로그 게이지의 중심을 추출하는데 적합하지 않기 때문에 최소자승법을 사용한다.

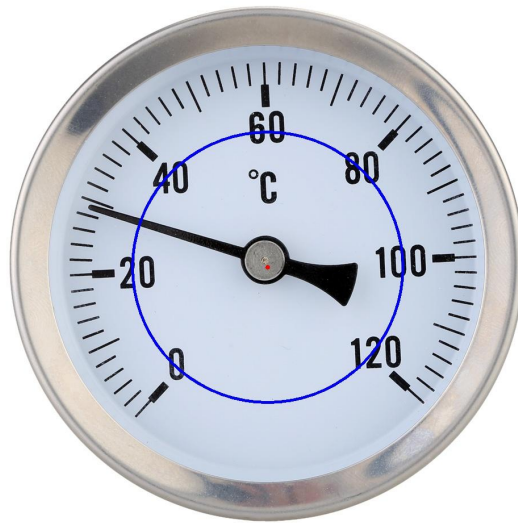


그림3.16 최소자승법을 이용한 중심 추출

4. 실험

4.1 실험 환경

실험 환경은 다음과 같다.

- 영상처리를 이용한 아날로그 게이지 정보 획득

OS : Window 10

CPU : intel i7-3630QM @2.40GHz

RAM : 12GB

- FCN을 이용한 아날로그 게이지 정보 획득

OS : Ubuntu 16.04

CPU : intel i5-6600 @3.30GHz

RAM : 8GB

GPU : Nvidia GTX 1080Ti

- Tesseract-OCR만 사용(FCN을 이용한 아날로그 게이지 정보 획득의 대조군)

OS : Window 10

CPU : intel i7-3630QM @2.40GHz

RAM : 12GB

4.2 영상처리를 이용한 아날로그 게이지 정보 획득 실험

총 4장의 이미지를 테스트에 사용하였다.

- 성공 : 게이지1
- 실패 : 게이지2, 게이지3, 게이지4



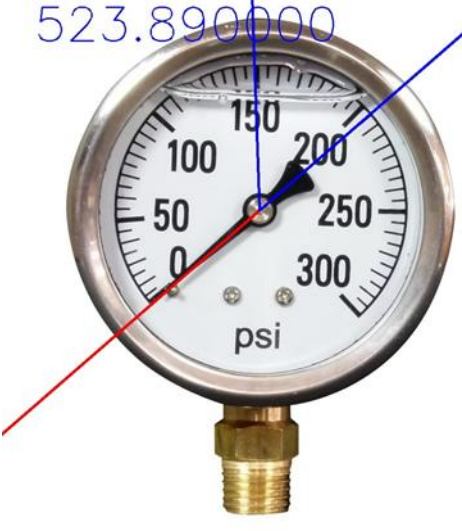


실패 원인 : 게이지 가운데의 원을 탐지하지 못 함(허프 변환 실패)

표 4.1 영상처리를 이용한 아날로그 게이지 정보 획득 실험 결과

	게이지1	게이지2	게이지3	게이지4
입력				
출력				

표 4.1은 영상처리를 이용한 아날로그 게이지 정보 획득 실험 결과를 나타낸다. 게이지1은 28.94의 값을 찾아내었으며 게이지2, 게이지3, 게이지4는 실패하였는데 원에 대한 허프 변환 과정이 게이지의 가운데 원을 찾지 못하였다. 그 이유는 전처리인 이진화 분할 과정에서 변수를 수동으로 결정하는데 이는 알고리즘이 대응할 수 있는 역량이 크게 줄어들었고 게이지의 가운데 원이 없거나 찾기 어려운 경우도 많기 때문이다. 전체 알고리즘의 다른 부분의 실험을 계속하기 위해 가운데 원을 수동으로 설정하여 진행하였다. 표4.2의 결과에 따르면 알고리즘이 전반적으로 불안정하다. 그 이유는 알고리즘에서 사용하는 모든 영상처리 방법의 변수는 사람이 수동으로 결정하는데 이는 다양한 환경을 가지는 아날로그 게이지에 적합하지 못하다는 의미이다.

표4.2 알고리즘의 추가 분석

결과	원인
	
<p>실패 원인 : 지침이 이진화 임계 방법에서 삭제</p>	
	<p>word: '150' word: ':00' word: '00' word: '50' word: '25C' word: '300'</p>
<p>실패 원인 : 정확한 숫자 판독 실패</p>	
	
<p>실패원인 : 해상도가 낮아서 숫자 인식 실패</p>	

4.3 FCN을 이용한 아날로그 게이지 정보 획득 실험

FCN의 결과를 보여주며 82장의 아날로그 게이지 이미지를 실험에 사용하였다.

표 4.3 Tesseract-OCR만 사용한 경우에 대한 결과

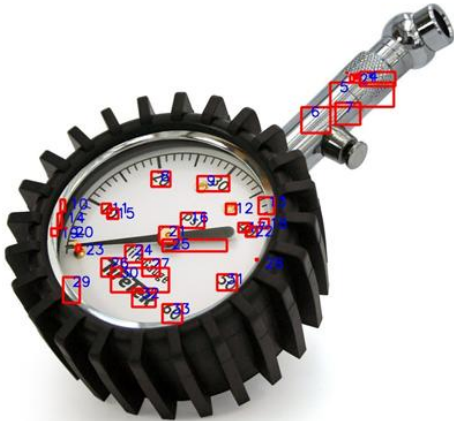
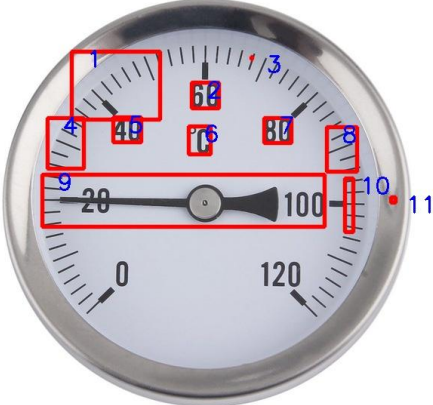
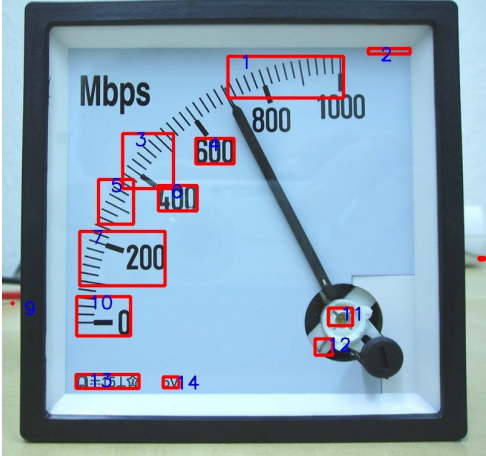
이미지 결과	관독 결과
	<pre> Box number :1 Text :' Box number :2 Text :'3 Box number :3 Text :" Box number :4 Text :“ Box number :5 Text ://’7«~r”” Box number :6 Text :[/2/ Box number :7 Text :f Box number :8 Text :’90 Box number :9 Text :0’70 Box number :10 Text :7 Box number :11 Text :1 Box number :12 Text :,’1” Box number :13 Text :’1? 33개중 13개만 표시 </pre>
	<pre> Box number :1 Text :y\\”\y\r Box number :2 Text :so Box number :3 Text :’ Box number :4 Text :\$ Box number :5 Text :40 Box number :6 Text :°c Box number :7 Text :80 Box number :8 Text :1; Box number :9 Text :Eyes-«100 Box number :10 Text :!; Box number :11 Text :~ </pre>
	<pre> Box number :1 Text :\\\\\\\\\\\\\\\\mi Box number :2 Text :w” Box number :3 Text :\\\\\\\\\\\\\\\\ Box number :4 Text :600 Box number :5 Text :\$ Box number :6 Text :mm Box number :7 Text :3200 Box number :8 Text :V Box number :9 Text :’ Box number :10 Text :Ean Box number :11 Text :V Box number :12 Text :,’1” Box number :13 Text :0:15;?! Box number :14 Text :5v </pre>

표 4.4 FCN을 이용한 아날로그 게이지 정보 획득


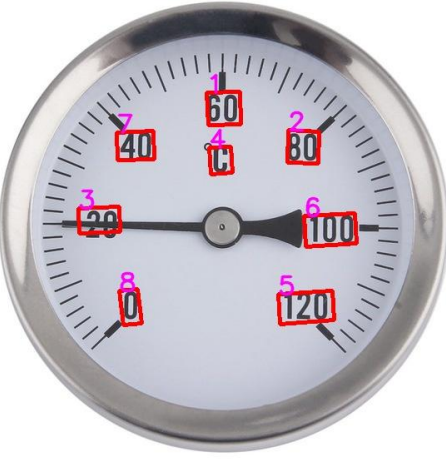
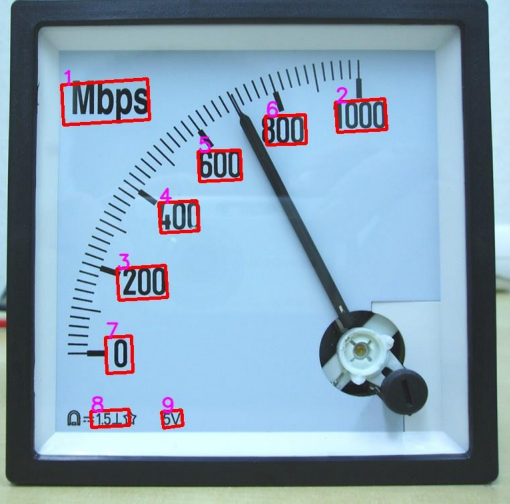
이미지 결과	판독 결과
	<p>Box numer :1 Text :W Box numer :2 Text :lirecau Box numer :3 Text :Psi Box numer :4 Text :40 Box numer :5 Text :60 Box numer :6 Text :30 Box numer :7 Text :10 Box numer :8 Text :20 Box numer :9 Text :501</p>
	<p>Box numer :1 Text :80 Box numer :2 Text :80 Box numer :3 Text :4R Box numer :4 Text :C Box numer :5 Text :120 Box numer :6 Text :100 Box numer :7 Text :40 Box numer :8 Text :0</p>
	<p>Box numer :1 Text :Mbps Box numer :2 Text :M Box numer :3 Text :700 Box numer :4 Text :hon Box numer :5 Text :m Box numer :6 Text :300 Box numer :7 Text : 0 Box numer :8 Text :M Box numer :9 Text :5V</p>

표 4.5 전체 테스트 데이터 셋에 대한 결과






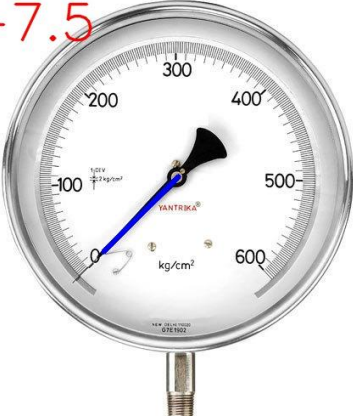
	정밀도	재현율	F1 score
Tesseract-OCR	11.78%	20.36%	13.48%
FCN + Tesseract-OCR	26.12%	35.21%	28.62%

정밀도는 모델이 정답이라 선택한 것 중 실제 정답의 비율이며, 정답 중에 모델이 선택한 것의 비율이다. F1 score는 정밀도와 재현율을 한번에 표현하는 방법이다. 아닌 것을 안 찾은 경우(True Negative)는 본 연구에서 판단 할 수 없으므로 정확도는 사용하지 않는다. 찾아야 하는 것을 찾은 경우(True Positive)는 Tesseract-OCR이 해당 숫자나 문자를 정확히 판단하였다면 해당하는 것으로 정했다. 찾아야 하는 문자는 아날로그 게이지의 내부에 존재하는 숫자와 문자로 한정하였다.

Tesseract-OCR을 단독으로 사용한 경우에 비해 정밀도가 2.2배, 재현율이 1.7배, F1 score가 2.1배 증가하였다. Tesseract-OCR의 주요 문제점은 숫자나 글자가 아닌 영역에서도 숫자나 글자라고 잘 못 검출하는 경우가 많았다. 또한 아날로그 게이지의 내부에 존재하는 숫자나 문자더라도 작은 노이즈로 인해 작은따옴표로 인식하는 경우가 많았다. 특히 눈금의 부분은 슬래쉬 문자로 오인하는 경우도 적지 않았다. 반면 FCN + Tesseract-OCR의 경우 FCN이 숫자나 문자의 영역을 지정해 줌으로써 Tesseract-OCR의 오탐지를 Tesseract-OCR만 사용할 때보다 낮추어 주었다. 다만 FCN도 놓치는 경우가 있고 정확한 위치를 못 찾을 경우 Tesseract-OCR이 오탐지하는 경우가 많았다.

표4.6은 FCN과 최소자승법이 사용된 알고리즘의 최종결과를 보여준다.

표4.6 FCN을 이용한 아날로그 게이지 정보 획득 실험 결과

<p>23.7</p>  <p>값: 23.7, 오차: 0.2</p>	<p>29.9</p>  <p>값: 29.6, 오차: 0.3</p>
<p>4.1</p>  <p>값: 4.1, 오차: 4.1</p>	<p>0.0</p>  <p>값: 0, 오차: 0</p>
<p>11.0</p>  <p>값: 11, 오차: 11</p>	<p>-7.5</p>  <p>값: -7.5, 오차: -7.5</p>

FCN을 이용한 아날로그 게이지 정보 획득 실험 결과는 아날로그 게이지의 정면이

미지만 사용하였으며 FCN과 Tesseract-OCR이 판별할 수 있는 이미지만 사용하였다.

5. 결론

본 논문에서는 영상처리 기법을 사용하여 아날로그 페이지의 정보를 획득하는 알고리즘과 FCN을 사용하여 아날로그 페이지의 정보를 획득하는 알고리즘을 제시하였다. 영상처리 기법을 사용하여 아날로그 페이지의 정보를 획득하는 알고리즘은 대체로 실패하였는데 그 이유는 알고리즘에 사용한 영상처리 기법 중 허프 변환, 이진화 임계 분할 방법은 문턱 값에 의해 동작하게 된다. 문턱 값은 사람이 직접 설정하게 되는데 이러한 방식은 다양한 환경에서 동작해야 하는 아날로그 페이지의 정보를 획득에 적합하지 못하였다. 반면 FCN을 사용하여 아날로그 페이지의 정보를 획득하는 알고리즘은 Tesseract-OCR을 단독으로 사용하는 경우보다 F1 score기준으로 2.1배 증가했으므로 아날로그 페이지의 정보를 획득하는데 적합하였다. 또한 ICDAR2015데이터와 직접 라벨링한 아날로그 페이지 데이터가 아날로그 페이지의 정보를 획득하는데 적합함을 보였다. 다만 오차가 많이 발생하는 경우가 있는데 글자를 중심으로 원을 찾은 점, 지침을 정확히 찾지 못 한점과 비율계산시 오차가 있다. FCN을 사용하면 사용하지 않은 경우보다 아날로그 페이지를 인식이 가능하나 오차가 크고 조건이 많으므로 이를 해결하기위해 Tesseract-OCR대신 Neural Network를 사용한 OCR을 추가할 것이며, 값을 찾기 위해 각도를 이용하지 않거나 좀 더 정확하게 찾을 수 있는 알고리즘을 연구할 것이다.

참 고 문 헌

- [1] 박관희, “스마트팩토리가 생산성과 품질 향상 불러온다”, 인더스트리 뉴스, <https://www.industrynews.co.kr>, 2017.12.18
- [2] J. Long, E. Shelhamer, and T. Darrell. “Fully convolutional networks for semantic segmentation.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3431- 3440, 2015
- [3] X. Liu, D. Liang, S. Yan, D. Chen, Y. Qiao, and J. Yan. Fots:Fast oriented text spotting with a unified network .arXivpreprint arXiv:1801.01671, 2018.
- [4] Rafael C. Gonzalez, Richard E. Woods, Digital Processing.(2003)
- [5] 오일석, “COMPUTER VISION”,(2014)
- [6] Richard O. Dudaz, Peter E. Use of the Hough transformation to detect lines and curves in pictures, CACM, vol.15, pp11-15.(1972)
- [7] R. Smith, “An Overview of the Tesseract OCR Engine” Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Nov 2007.
- [8] 손기석, 조강현, “아날로그 게이지 정보 해석” Institute of Control, Robotics and Systems(ICROS2018), 2018.
- [9] 손기석, 양재완, 조강현, “산업용 아날로그 게이지 숫자 인식” Institute of Control, Robotics and Systems(ICROS2019), 2019.
- [10] “International Conference on Document Analysis and Recognition(ICDAR)”, <http://rrc.cvc.uab.es/>, 2015.
- [11] Rosenblatt, F. (1958) The perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev., 65, 386408.
- [12] Nair and G. E. Hinton. Rectified linear units improve restrictedbo ltzmann machines. InICML, 2010.