

COMPUTER VISION



DEEP
LEARNING



컴퓨터 비전과 딥러닝

Chapter 01

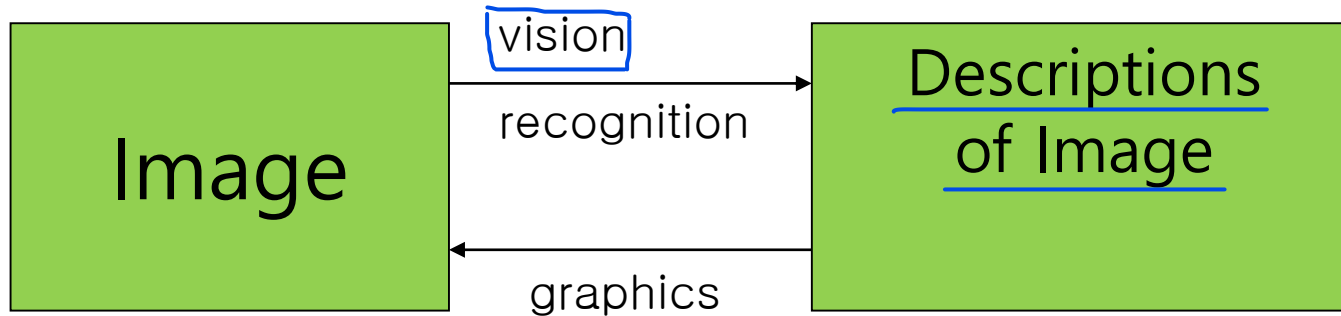
주교재: 한빛아카데미(2023)

인간의 시각을 흉내 내는 컴퓨터 비전

엄 종 석, jsum@hansung.ac.kr

eclass : Q&A

컴퓨터비전이란?



Forsyth and Ponce

- Extracting descriptions of the world from picture or sequences of pictures.

Ballard and Brown

- The construction of explicit, meaningful description of physical objects from images

컴퓨터비전이란?

- Make machines to see, or to understand/interpret the scenes (image and videos) like human do.
 - Measure the properties of 3D world using visual data
 - build computational models for human visual system
 - build autonomous systems to perform some of the tasks which the human visual system can perform and even surpass it in many cases
 - Recognize objects, activities, relations, and context
- Vision is difficult because it is an inverse problem.
 - Understanding 3D world from 2D image.
 - Many interpretations are possible (ill-posed problem).

왜 컴퓨터 비전인가?

- Vision is how we experience the world.
- An image is worth 1000 words.
- Images and movies are everywhere.
- Fast-growing collection of useful applications
 - automated surveillance
 - object detection, tracking & recognition
 - gesture recognition
 - intelligent robots, stereo vision, 3D Reconstruction
 - autonomous vehicle
- Many biological system rely on vision
 - If the eye can do it, so can the machine.
- Greater understanding of human vision
- Powerful algorithms for object recognition
- Camera and computing resources are cheap.



왜 컴퓨터 비전인가?

■ 팡창하는 응용

- 오락(기능성 게임, 사진 변형), 교통(지능형 교통관리, 자율 주행), 보안(생체인식), 산업(제품 검사, 조립, 물류), 계산 사진학(파노라마 영상, 영상 합성), 의료(의료 영상 진단, 수술용 로봇), 과학(전자현미경 연상 분석), 농업, 군사, 모바일(증강현실 스마트폰 앱, 구글 렌즈)



(a) 구글 자동차(자율 주행)



(b) 수술용 로봇 다빈치



(c) 화성 탐사선



circuit board inspection



(d) 딸기 따는 로봇



(e) 군사용 로봇 빅독



(f) 리프스냅



구글 렌즈

그림 1-3 컴퓨터 비전의 응용 사례

왜 컴퓨터 비전인가?

■ 팽창하는 응용

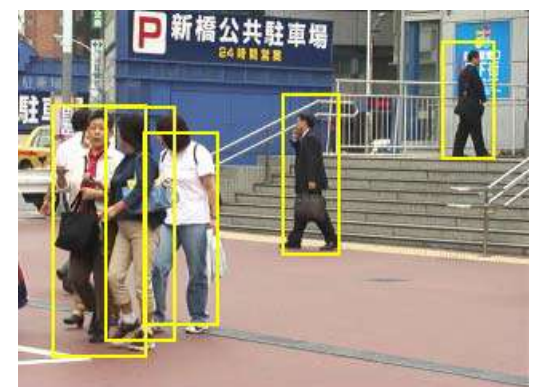
- 오락(기능성 게임, 사진 변형), 교통(지능형 교통관리, 자율 주행), 보안(생체인식), 산업(제품 검사, 조립, 물류), 계산 사진학(파노라마 영상, 영상 합성), 의료(의료 영상 진단, 수술용 로봇), 과학(전자현미경 영상 분석), 농업, 군사, 모바일(증강현실 스마트폰 앱, 구글 렌즈)



ADAS(Advanced Driver Assistance System)



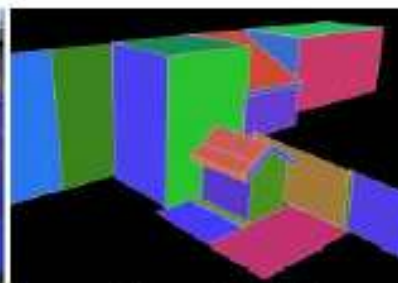
Object detection and Classification



Input Photographs



2D Sketching Interface



Geometric Model



Texture-mapped model

3D Modeling

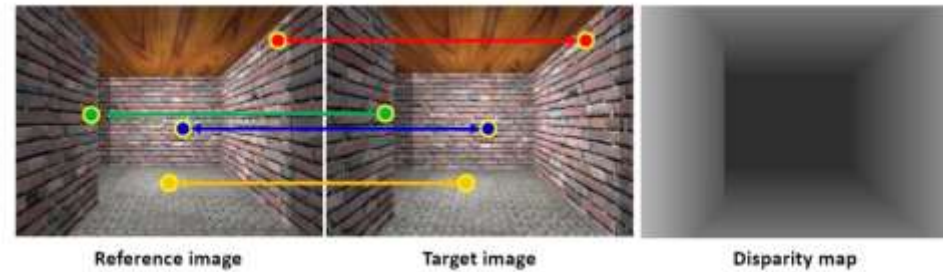
왜 컴퓨터 비전인가?

■ 팡창하는 응용

- 오락(기능성 게임, 사진 변형), 교통(지능형 교통관리, 자율 주행), 보안(생체인식), 산업(제품 검사, 조립, 물류), 계산 사진학(파노라마 영상, 영상 합성), 의료(의료 영상 진단, 수술용 로봇), 과학(전자현미경 영상 분석), 농업, 군사, 모바일(증강현실 스마트폰 앱, 구글 렌즈)



Automatic panoramas



3D Reconstruction

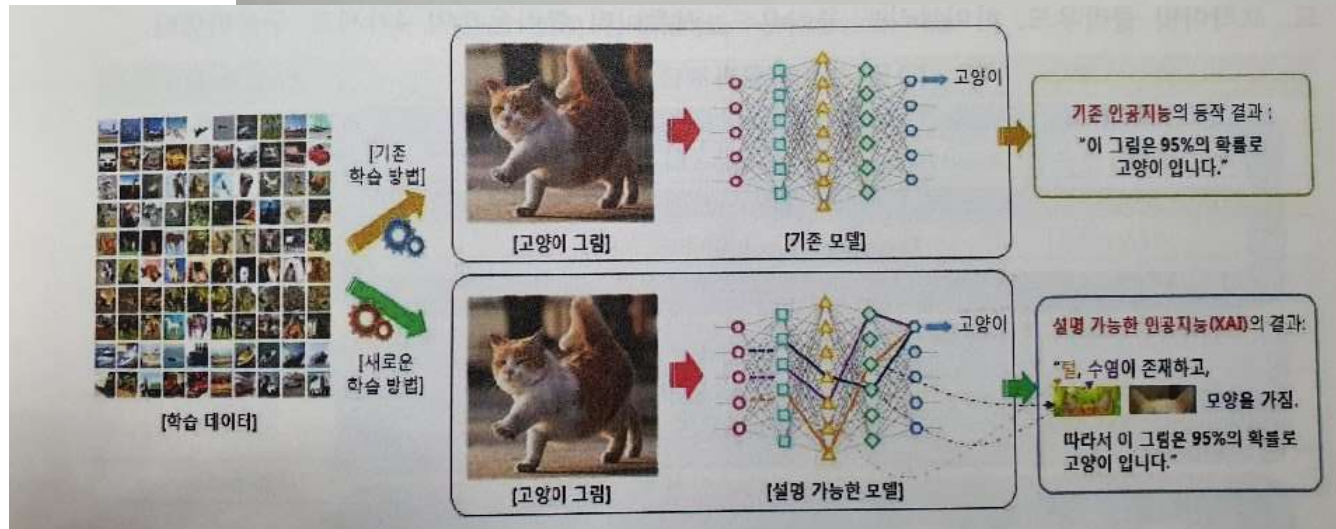


의료 영상 진단

왜 컴퓨터 비전인가?

■ 팡창하는 응용

- 시각 이해
- XAI



<https://www.microsoft.com/en-us/ai/seeing-ai>

왜 컴퓨터 비전인가?

■ 팡창하는 응용

- 생성 비전: GAN(Generative Adversarial Network)



OpenAI 사의 생성형 AI 서비스 DALL·E2 사용하여 다음 입 문장 입력으로 “무한한 우주, 신스웨이브, 디지털 아트에서 화성의 카메라 쪽으로 걸어가는 운동 여성의 몸을 가진 여성 우주 비행사의 아래에서 광각 촬영” 이미지 생성(2022)

출처: [\[해외 크리에이티브\] AI가 디자인한 코스모폴리탄 표지 - 매드타임스\(MADTimes\)](#)

[Which Face Is Real?](https://www.whichfaceisreal.com)

<https://www.whichfaceisreal.com>



그림 1-9 Which face is real? (왼쪽이 진짜)

- 이미지 생성형 AI 서비스
 - [Adobe: Firefly](#)
 - Discord messenger: Midjourney
 - OpenAI: DALL-E3

과학적 접근과 공학적 접근

■ 컴퓨터 비전의 목표

- 목표 1 : 사람의 시각에 맞먹는 인공 시각을 만든다.



(a) 목표1 - 제약이 없는 상황

- 목표 2 : 한정된 범위에서 특정한 임무를 달성하는 인공 시각을 만든다.



(b) 목표2 - 제약이 있는 상황

그림 1-4 컴퓨터 비전의 목표

과학적 접근과 공학적 접근

■ 과학적 접근

- 목표 1을 달성하려는 노력
- 사람 시각의 원리를 밝혀낸 다음 컴퓨터로 모방할 수 있어야 가능
- 뇌 과학의 주요 관심사
- 상황을 인식하기 위해서는 지식 표현, 학습, 추론, 창작 같은 인공 지능이 필수

■ 어려운 이유

- 역 문제(2차원 영상으로 3차원 세계를 추론, 물체와의 거리 정보가 없음)
- 불량 문제(문제의 답이 유일하지 않음, 방정식의 수 < 미지수의 수)
- 다양한 변형 발생 (기하학적 변환, 광도 변환, 잡음)

과학적 접근과 공학적 접근

■ 공학적 접근

- 목표 2를 달성하려는 노력
 - 특정 상황에서 특정 임무를 수행하는 실용 시스템 구축
- 성공적인 시스템
 - 많은 응용 현장에서 쓰고 있음
 - 컴퓨터 비전이 사람보다 뛰어난 경우
 - 예, 엔진 실린더 지름의 정밀 측정, 칩 검사, IBM Watson 암 진단율

■ 실용적인 성능 달성의 어려움

- 여전히 역 문제, 불량 문제, 다양한 변환 발생
- 영상은 숫자 배열 형태 => 이에 근거한 인식은 여전히 어려운 문제

125 134 125 122 127 120 130 139 135 139 140 133 127 127 130 139 135 138 133 137 136 134 130 125 121
117 129 114 116 120 122 118 120 122 117 122 126 124 117 106 100 99 102 105 120 116 113 108 105 106 111
109 110 105 102 112 125 130 135 147 171 191 184 189 174 157 139 124 107 90 92 87 88 92 90 88 88
108 105 100 116 117 129 163 195 210 217 205 215 211 136 195 176 167 143 117 51 80 77 38 58 84 79
107 103 102 120 145 173 200 193 172 195 138 141 136 123 118 125 139 143 137 121 99 84 85 88 82 81
104 107 115 134 159 171 170 136 115 129 107 83 85 82 80 83 90 103 113 125 108 93 91 90 86 83
107 120 137 160 150 125 139 150 167 174 115 99 94 90 98 99 89 87 91 104 103 99 97 95 94 95
111 133 196 134 151 197 189 206 216 212 136 114 92 83 97 110 108 100 98 97 101 101 96 92 103 120
130 145 164 165 185 213 219 210 212 196 158 108 123 137 137 123 111 121 134 145 132 130 147 159 163 171
138 151 170 106 135 215 222 211 214 210 209 160 152 151 157 163 166 167 166 159 195 180 180 130 135 193
142 153 171 190 190 204 218 213 207 214 218 213 204 195 192 189 183 176 173 161 159 163 171 183 189 187
141 151 164 188 178 180 197 204 201 197 196 196 193 190 187 178 163 157 156 156 161 163 166 174 186 192
144 151 160 185 183 176 176 187 182 181 188 193 184 178 177 174 165 156 151 148 163 177 182 188 200 200
152 160 168 178 193 193 182 180 180 174 172 164 161 159 154 148 140 143 149 173 164 130 190 190 199 205
159 168 178 178 202 206 197 194 187 176 176 187 172 179 180 176 176 188 203 215 212 206 204 202 204 205
161 171 185 197 210 204 189 211 210 208 212 219 210 206 215 225 226 220 215 214 209 210 214 216 211 200

그림 1-5 숫자 배열 형태의 영상



문제 해결 도구

■ 자료 구조와 알고리즘

- 처리과정에서 발생하는 데이터 표현: 배열, 트리, 그래프, 힙, 탐색 트리 등
- Optimization algorithms: 탐색 방법, 동적 프로그래밍, 한정 분기 등
- 고속 처리가 주된 관심

■ 수학

- 선형 대수, 미적분학(변화량, 극점 탐색), 확률과 통계(의사결정, 분류) 등
- 최적화 문제 풀이(비용함수 정의, 비용을 최소화 시키는 매개변수를 구하는 문제)

■ 기계 학습

- 기계 학습을 활용하는 사례 급증
- 신경망(DNN, CNN), SVM, 에이더부스트(AdaBoost), 임의 숲(random forests) 등

성능 평가

인식 성능 측정

$$\text{정인식률} = \frac{c}{N}, \text{기각률} = \frac{r}{N}, \text{오류율} = \frac{e}{N}$$

이때 c = 맞는 샘플수, r = 기각한 샘플수, e = 틀린 샘플수
($N = c + r + e$)

기각은 분류를
포기한 경우
(1.1)

- 부류(class)가 심한 불균형일 때 부적절
(예, 칩 검사에서 불량률이 0.1%라면, 임의의 짐작 분류기의 정인식률은 99.9%)

혼동 행렬(Confusion matrix)

- 오류 경향을 세밀하게 분석하는데 사용

표 1-1 부류가 두 개인 경우의 혼동 행렬

참부류 \ 분류 결과	ω_1	ω_2
	ω_1	ω_2
ω_1	$n_{11}(\text{TP})$	$n_{12}(\text{FN})$
ω_2	$n_{21}(\text{FP})$	$n_{22}(\text{TN})$

성능 평가

- 참/거짓 긍정률, 참/거짓 부정률, 재현률(recall)과 정확률(precision), F 측정

$$\begin{aligned}\text{거짓 긍정률}\left(\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}\right) &= \frac{n_{21}}{(n_{21} + n_{22})} \\ \text{거짓 부정률}\left(\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}\right) &= \frac{n_{12}}{(n_{11} + n_{12})} \\ \text{참 긍정률}\left(\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}\right) &= \frac{n_{11}}{(n_{11} + n_{12})} \\ \text{참 부정률}\left(\text{TNR} = \frac{\text{TN}}{\text{FP} + \text{TN}}\right) &= \frac{n_{22}}{(n_{21} + n_{22})}\end{aligned}\tag{1.2}$$

$$\begin{aligned}\text{정확률} &= \frac{n_{11}}{(n_{11} + n_{21})} \\ \text{재현율} &= \frac{n_{11}}{(n_{11} + n_{12})}\end{aligned}\tag{1.3}$$

$$\begin{aligned}F_{\beta} &= (1 + \beta^2) \frac{\text{정확률} \times \text{재현율}}{\beta^2 \times \text{정확률} + \text{재현율}} \\ F_1 &= \frac{2 \times \text{정확률} \times \text{재현율}}{\text{정확률} + \text{재현율}}\end{aligned}\tag{1.4}$$

성능 평가

예제 1-1 얼굴 검출의 성능 측정

[그림 1-10]은 세 개의 영상을 가진 데이터베이스에서 얼굴을 검출한 결과를 보여준다. 앞에서 다룬 정확률, 재현율, F_1 측정치를 구해보자.

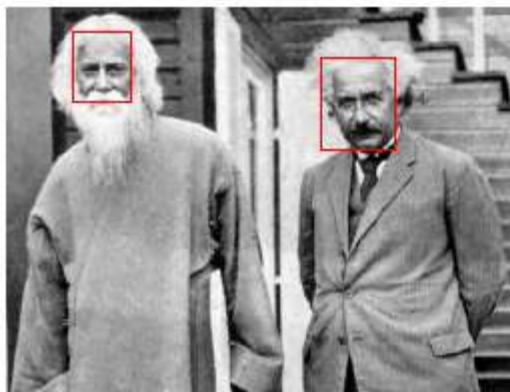
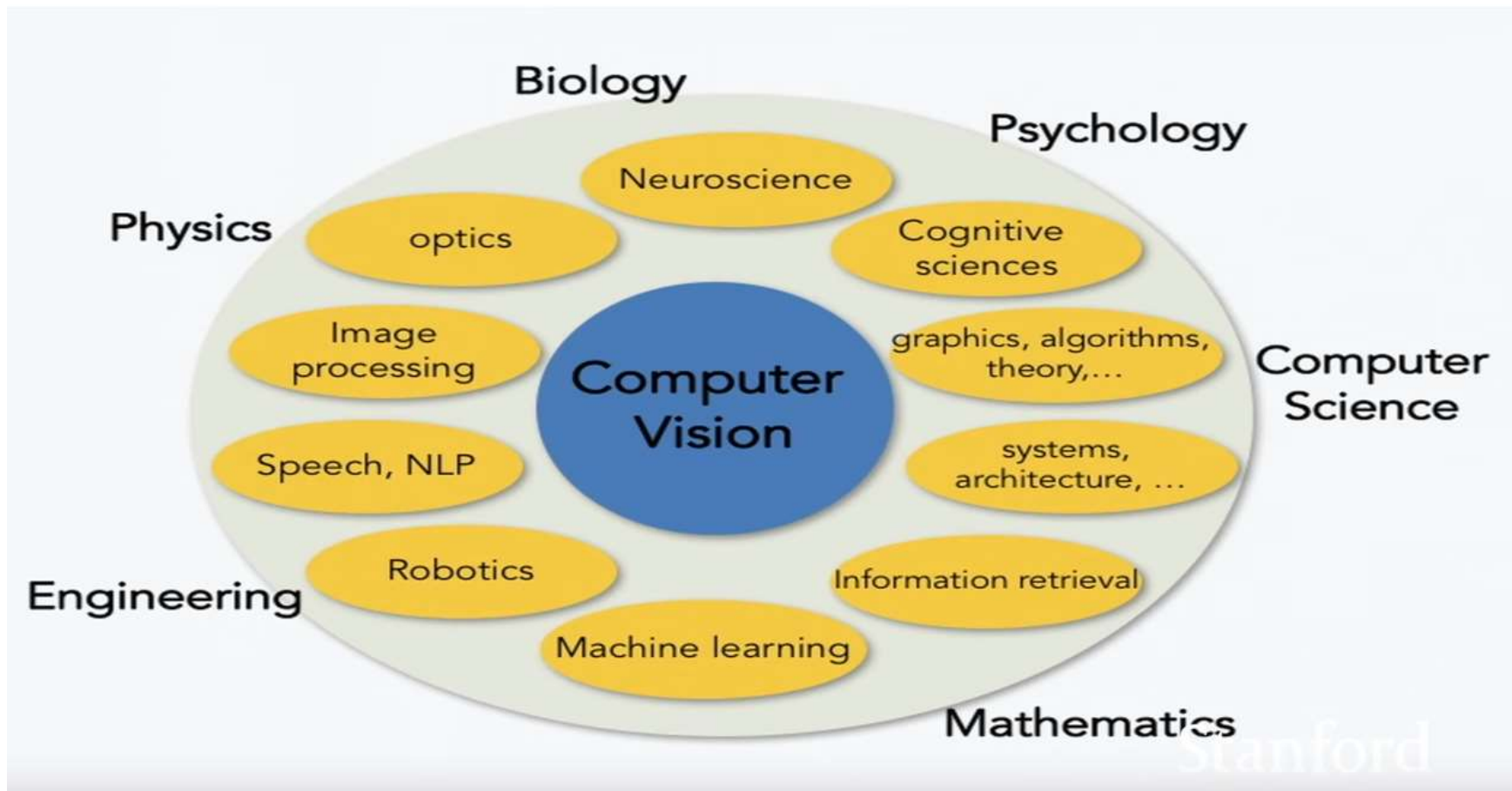


그림 1-10 얼굴 검출 성능

총 15개의 얼굴 중 12개를 옳게 검출했으므로 참 긍정 $n_{11}=12$, 세 개의 얼굴을 못 찾았으므로 거짓 부정 $n_{12}=3$, 그리고 얼굴 아닌 곳을 얼굴로 검출한 것이 두 개이므로 거짓 긍정 $n_{21}=2$ 이다. 따라서 정확률은 $12/14$ 이고 재현율은 $12/15$ 이다. F_1 측정은 $24/29$ 이다.

인접 학문

- 상호 협력이 강해지는 추세



학습을 위한 자원

■ 웹 사이트

- CVonline(<http://homepages.inf.ed.ac.uk/rbf/CVonline/>)
 - 컴퓨터 비전의 주제를 계층적으로 나누고, 그들 주제에 대한 설명을 위키피디아를 이용하여 제공한다.
 - 데이터베이스, 문헌, 소프트웨어, 교육을 위한 최신 자료도 풍부하다.
- VisionBib.Com(<http://www.visionbib.com/index.php>)
 - 컴퓨터 비전과 관련된 논문을 망라하여 제공한다.
 - 주제어, 저자, 시기, 학술지, 학술대회에 따라 검색이 가능하다.
- Computer Vision Online(<http://www.computervisiononline.com/>)
 - 소프트웨어, 데이터베이스, 그리고 책에 대한 최신 정보를 제공한다.
 - 컴퓨터 비전 분야에서 주목할 만한 뉴스도 알려준다.

■ 데이터베이스

- <http://host.robots.ox.ac.uk/pascal/VOC/index.html>
- <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>
- PASCAL Visual Object Challenge(2006~2012): <http://host.robots.ox.ac.uk/pascal/VOC/>
- ImageNet Large Scale Visual Recognition Challenge(ILSVRC): <http://www.image-net.org/challenges/LSVRC/>

학습을 위한 자원

PASCAL Visual Object Challenge (20 object categories) [Everingham et al. 2006-2012]

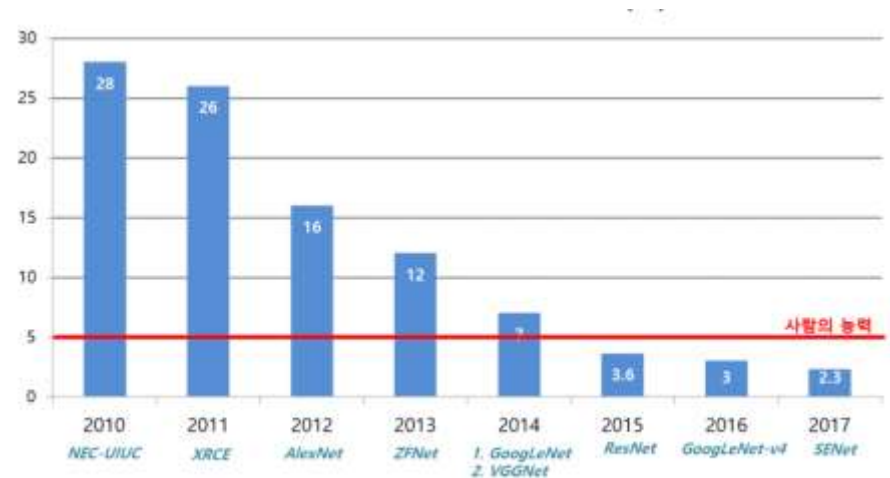
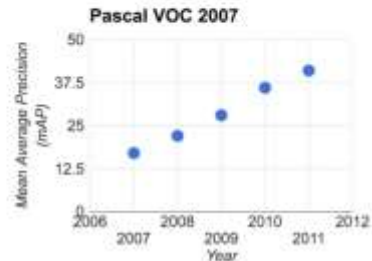
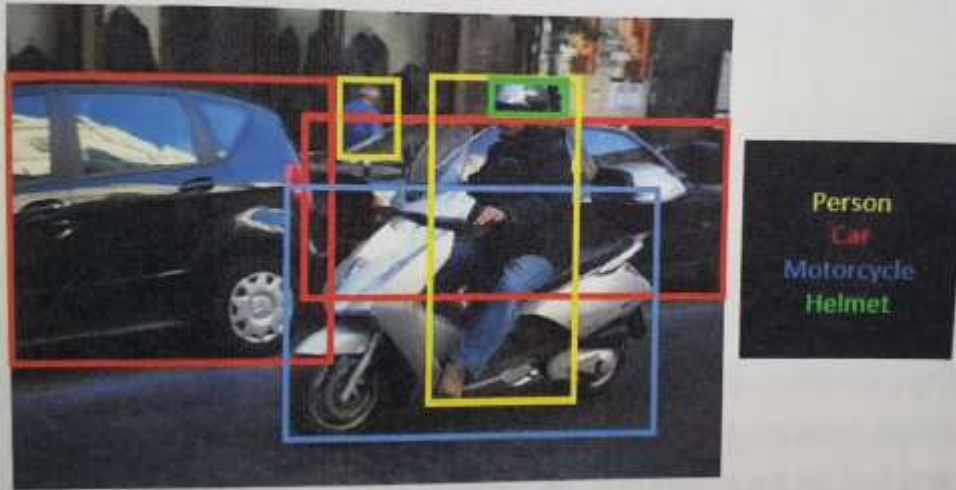


그림1. ILSVRC 대회 역대 우승 알고리즘들과 인식 에러율.

2012 CNN 최초 적용

[그림 I-25] ImageNet 경진대회 시각 인식 결과



ImageNet: 1400백만 가량의 영상과 21,841개의 category로 labeling

ILSVRC: ImageNet Large Scale Visual Recognition Challenge (2010~2017)

학습을 위한 자원

■ 학술대회

- IEEE International Conference on Computer Vision(ICCV)
- IEEE Internal Conference on Computer Vision and Pattern Recognition(CVPR)
- ACM SIGGRAPH

■ 학술지

- IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI)
- International Journal on Computer Vision(IJCV)
- Image and Vision Computing

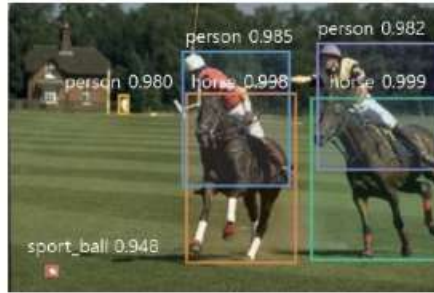
컴퓨터 비전이 풀어야 할 문제

■ 기본 문제

- 분류
- 검출
- 분할
- 추적
- 행동 분석
- ...



(a) 분류



(b) 검출



(c) 분할



(d) 추적(<https://motchallenge.net/vis/MOT17-09-SDP>)



(e) 행동 분석(<https://github.com/mostafa-saad/deep-activity-rec#dataset>)

그림 1-13 컴퓨터 비전이 풀어야 할 문제

컴퓨터 비전 알고리즘과 프로그래밍

■ 고전 컴퓨터 비전과 딥러닝 컴퓨터 비전

- 대략 2010년을 기점으로 방법론의 대전환. 고전 방법은 규칙 기반, 딥러닝은 데이터 중심
- 이 책의 4~6장은 고전 알고리즘, 7~13장은 딥러닝 위주
- 둘 다 이해하는 것이 중요

■ 프로그래밍 언어

- C/C++, 자바, 파이썬을 주로 사용. 이 강의에서는 파이썬을 사용
- 주로 C/C++는 알고리즘 구현, 파이썬은 인터페이스 언어로 활용됨
- 컴퓨터 비전 지원하는 OpenCV와 딥러닝 지원하는 텐서플로 라이브러리 사용

■ 파이썬

- 배열 처리에 유리
- 파이썬 기초를 다지고 이 책을 공부해야 함

[C 언어]

```
for(i=0; i<n; i++) z[i]=x[i]+y[i];
```

[파이썬 언어]

```
z=x+y
```

교재

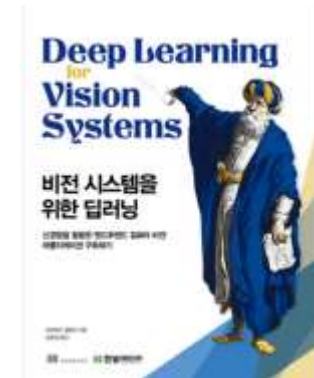
■ 주교재

- 컴퓨터 비전과 딥러닝
 - 한빛아카데미(2023), 오일석 지음



■ 참고문헌

- 1. OpenCV 4로 배우는 컴퓨터 비전과 머신 러닝
- 2. 비전 시스템을 위한 딥러닝 : 한빛아카데미
- 3. <http://opencv.org>
 - Documentation, Tutorials
- 4. Computer Vision – Algorithms and Applications
 - second edition
 - Springer: Richard Szeliski
 - 최신 컴퓨터 비전 내용을 방대하게 담음
 - 무료 pdf 제공 (<https://szeliski.org/Book>)



■ 파이썬 프로그래밍 기초

- 이 책의 부록 A
- 온라인 학습
 - <https://learnpython.org>
 - 코세라 [Top Python Courses - Learn Python Online \(coursera.org\)](https://www.coursera.org/learn/python)



무엇을 다루는가

학습 준비	<ul style="list-style-type: none">1장 인간의 시각을 흉내 내는 컴퓨터 비전2장 OpenCV로 시작하는 컴퓨터 비전3장 영상 처리
고전 컴퓨터 비전	<ul style="list-style-type: none">4장 에지와 영역5장 지역 특징6장 비전 에이전트
딥러닝 컴퓨터 비전	<ul style="list-style-type: none">7장 딥러닝 비전8장 컨볼루션 신경망9장 인식10장 동적 비전11장 비전 트랜스포머12장 3차원 비전13장 생성 비전

강의계획

■ 1주: 컴퓨터 비전 소개

- Anaconda 설치 및 기본사용법(파이썬 기초 참고)

■ 2주: 영상처리 기본: 히스토그램, 이진영상 – 컬러영상

- 화소 값에 접근(영상 조회), 간단한 영상 산술 실행, 컬러 공간 변환, 사용자 인터페이스

■ 3주: 영상처리 – 점 연산, 영역 연산, 기하연산, 모폴로지

- 히스토그램 평활화, 컨볼루션, 보간, 형태학적 필터를 이용한 처리

■ 4주: 에지 검출

- 에지 검출 기초 - 케니 에지, 허프 변환

■ 5주: 영역 분할 – 물체 분할, 영역 특징

- GrabCut 알고리즘을 이용한 전경 객체 추출(영상 분할)

■ 6주: 지역 특징 검출

- 지역 특징 검출의 기초 - 해리스 코너 검출

■ 7주: 스케일 불변 특징점 검출

- 스케일 공간, 크기에 불변한 특징 검출(SIFT:Scale Invariant feature Transf.)

■ 8주: 중간고사

강의계획

- 9주: 매칭 기초 - 매칭 알고리즘, 기계학습
 - 매칭(FLANN: 빠른 최근점 이웃 탐색), 호모그래피 추정(RANSAC)
- 10주: 비전에이전트:
 - 매칭 응용(교통 약자 보호구역 인식), 파노라마 영상제작
- 11주: 딥러닝 비전(DNN)
 - 학습 알고리즘, 다층퍼셉트론 구현하기(MNIST), 우편번호 인식기 v_1
- 12주: 컨볼루션 신경망(CNN)
 - 학습 알고리즘 및 구현, LeNet-5, 전이학습, 우편번호 인식기 v_2
- 13주: 인식(YOLO v3):
 - 물체 검출하기, 물체 분할하기
- 14주: 동적비전
 - 비디오 시퀀스 처리 -Lucas-Kanade 추적 알고리즘
 - 물체 추적 - 이동 물체 탐지
- 15주: 기말고사

grades

- 1. Midterm exam: 40%
- 2. Final exam: 40%
- 4. Homeworks 10%
- 5. Attendance 10%(-1.5/1시간(75분 수업) 결석)