

自己教師あり学習技術マップ

作成: 東京工業大学 井上中順

2020年以前

初期の自己教師あり学習

画像の回転角などの疑似ラベルを用いた学習方法が提案され、教師なしデータから画像表現の学習が可能であることが示された。

Pretext tasks

Rotation

S. Gidaris+, ICLR18

Colorization

R. Zhang+, ECCV16

Jigsaw

M. Noroozi+, ECCV16

2020~2022年

対照学習の発展

SimCLR を筆頭に Noise Contrastive Estimation に基づいた学習方法が提案された。

この頃から、ImageNet データセットにおいて、自己教師あり学習モデルが教師あり学習モデルの画像認識精度を上回り始め注目を集めることとなる。

Contrastive learning

SwAV

M. Caron+, NeurIPS20

BYOL

J.-B. Grill+, NeurIPS20

SimCLR v2

T. Chen+, 2020

SimCLR

T. Chen+, ICML20

MoCo v3

X. Chen+, ICCV21

SimSiam

X. Chen+, CVPR21

MoCo v2

X. Chen+, 2020

MoCo

K. He+, CVPR20

DINOv2

M. Oquab+, TMLR (24)

DINO

M. Caron+, ICCV21

VICReg

A. Bardes+, ICLR22

Barlow Twins

J. Zbontar+, ICML21

Masked image modeling

2023年~現在

マスク画像モデリング (MIM)

MAE を筆頭にマスクを用いた学習方法が確立されつつある。MIM はマスク言語モデルの画像版と位置付けられるもので、現在でも対照学習の損失関数との組合せや、特徴量をターゲットとした数多くの手法が提案されている。

Token target

BEiT3

W. Wang+, CVPR23

BEiT2

Z. Peng+, 2022

BEiT

H. Bao+, ICLR22

iBoT

J. Zhou+, ICLR22

MAE

K. He+, CVPR22

EsViT

C. Li+, ICLR22

SdAE

Y. Chen+, ECCV22

GreenMIM

L. Huang+, NeurIPS22

SemMAE

G. Li+, NeurIPS22

MixedAE

K. Chen+, CVPR23

Feature target

A2MIM

S. Li+, ICLR23

MSN

M. Assran+, ECCV22

LocalMIM

H. Wang+, CVPR23

BootMAE

X. Dong+, ECCV22

ccMIM

S. Zhang+, ICLR23

RobustMAE

Q. Huang+, ICCV23

R-MAE

D.-K. Nguyen+, ICLR24

MAE-Lite

S. Wang+, ICML23

HPM

H. Wang+, CVPR2023

SparseMAE

A. Zhou+, ICCV23

MFM

Y. Liu+, ICCV23

SIM

C. Tao+, CVPR23

VisualAtom

S. Takashima+, CVPR23

OFDB

R. Nakamura+, ICCV23

SegRCDB

R. Shinoda+, ICCV23

FractalDB

H. Kataoka+, ACCV20

RCDB

H. Kataoka+, CVPR22

PC-FractalDB

R. Yamada+, CVPR22

数式駆動教師あり学習

自然画像を一切用いない学習方法も存在する。

数式駆動教師あり学習 (FDSL) はその最たる例であり、最近の研究では ImageNet に匹敵する事前学習効果が報告されている。

Formula-driven supervised learning

今後の展開

画像と言語を横断したマルチモーダル基盤モデルが急速な発展を遂げている。本技術マップでは静止画像に対する自己教師あり学習を俯瞰したが、3D データや映像データに対する MIM も成功例が多く、多種多様な形式の視覚データが統一的に扱える日も近くなりそうだ。