

prerequisite:

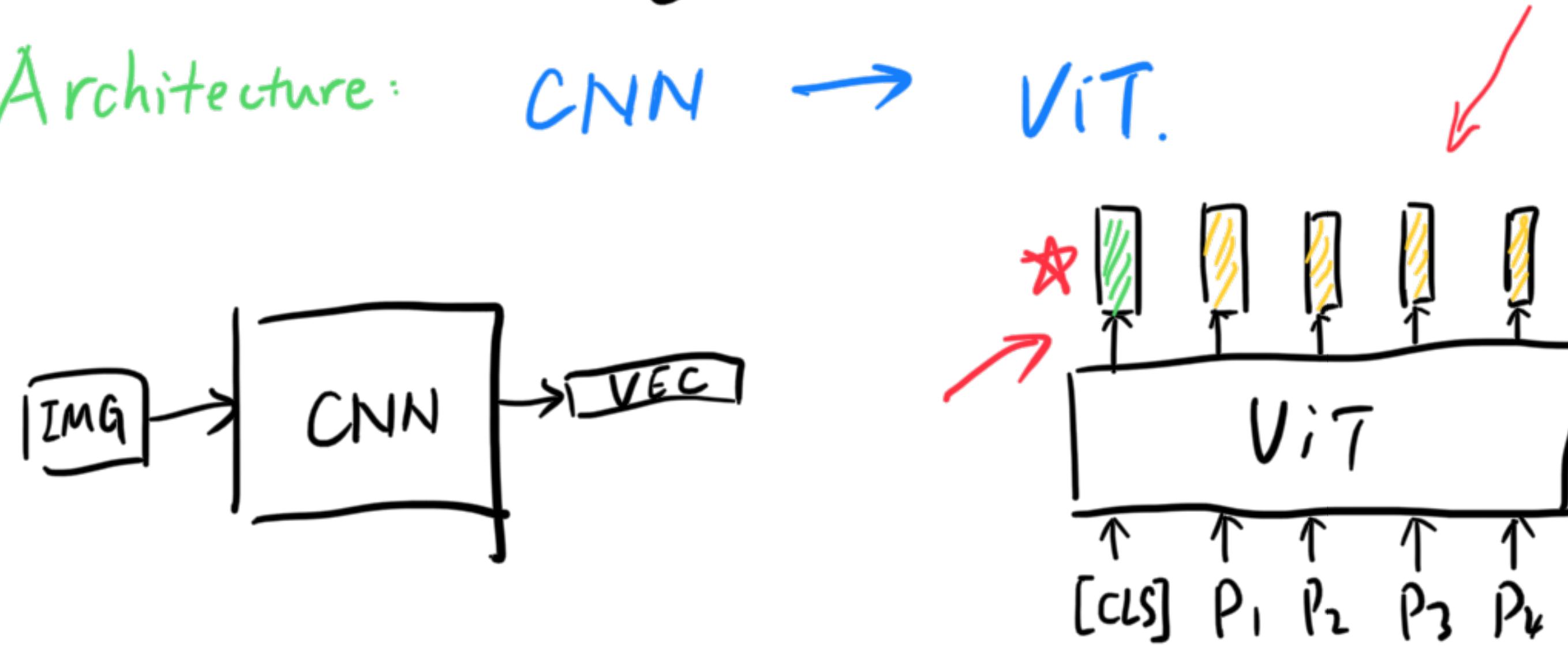
- \* ViT
- \* Self-supervision \*
- \* VAE → CogView



BERT

Contrastive Learning → Denoise Auto-Encoding.

Architecture: CNN → ViT.



Performance: Linear Probing → Fine-tuning

Background

methods

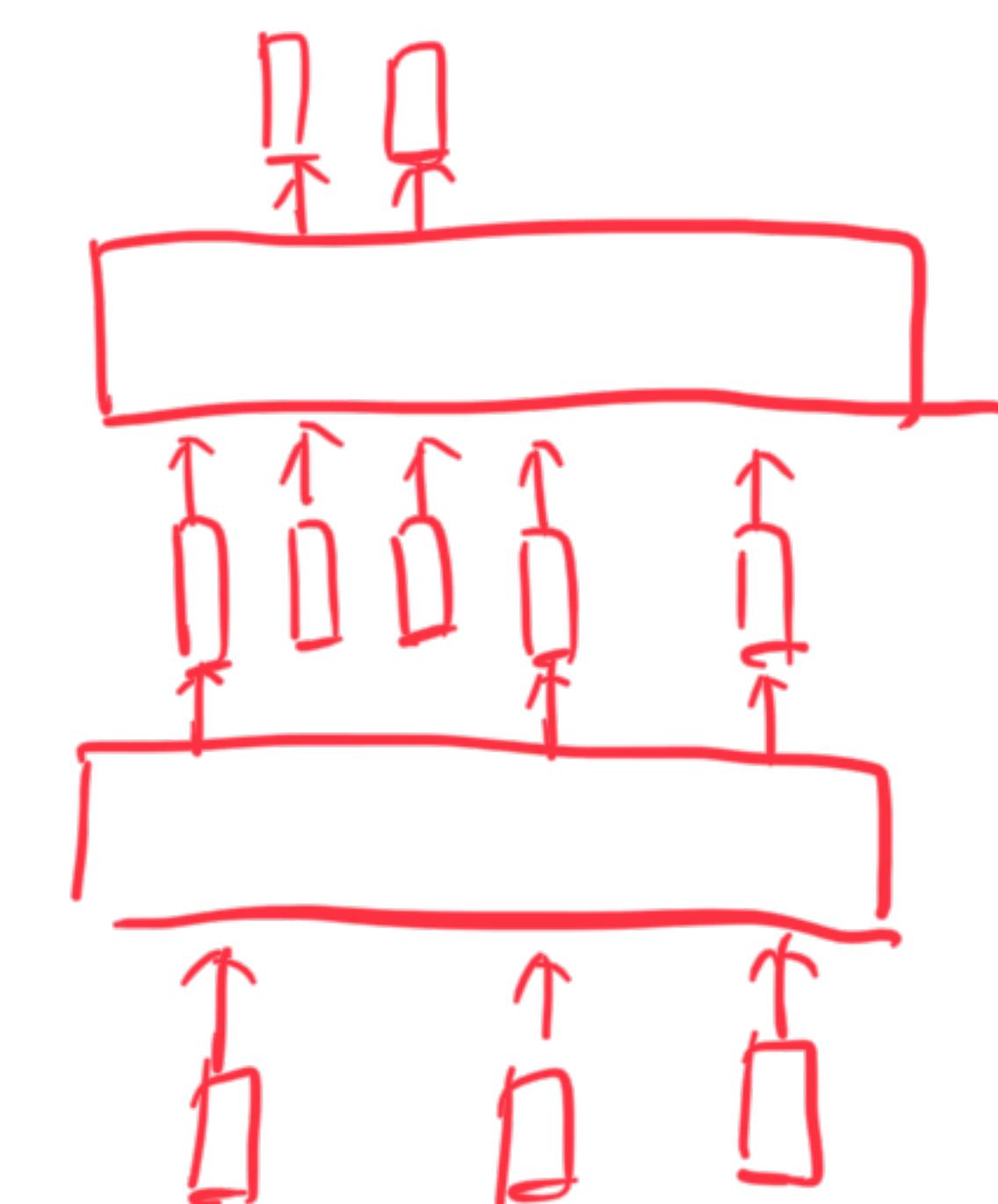
masked patch prediction → BERT + regression

BEiT → BERT + dVAE

SiT → DAE + CL

MAE → BART

iBOT → BERT + BYOL/DINO.



BERT 15%

MAE 75%

future direction → pretext task design (BERT, BART, etc.)

→ multi-modality (unified training)

→ decoder application (prompt?)

标题	创建者	日期
► BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Abstractive Summarization	Lewis 等。	2019-10-29
► iBOT: Image BERT Pre-Training with Online Tokenizer	Zhou 等。	2021-11-15
► BEiT: BERT Pre-Training of Image Transformers	Bao 等。	2021-06-15
► Masked Autoencoders Are Scalable Vision Learners	He 等。	2021-11-11
► SiT: Self-supervised vision Transformer	Atito	
► An Empirical Study of Training Self-Supervised Visual Transformers	Chen 等。	2021-04-05