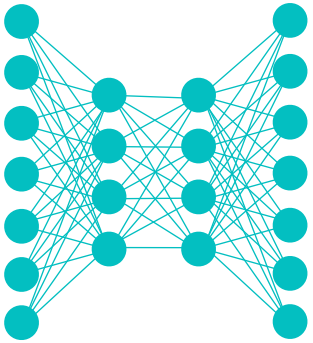
Lecture Notes for

Neural Networks and Machine Learning

Final Transformers

Self-Supervision

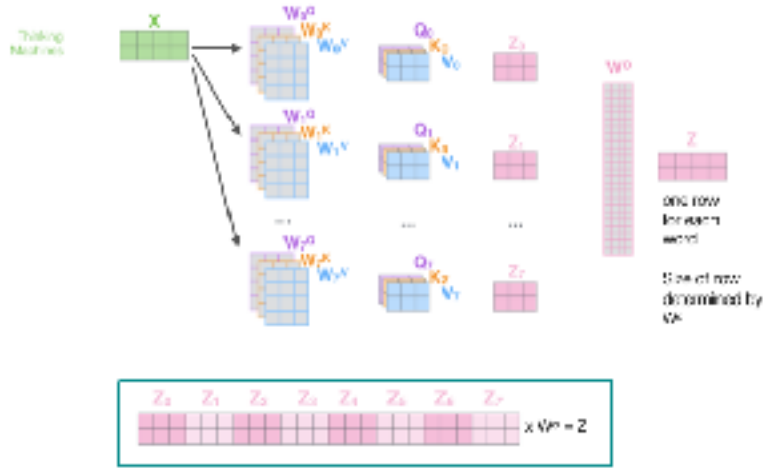


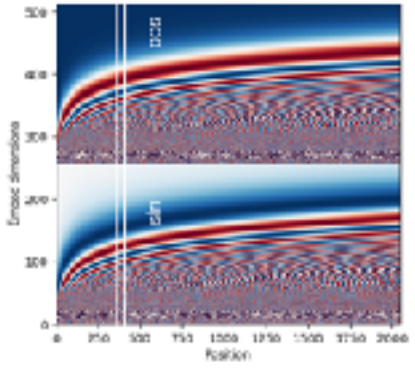
Logistics and Agenda

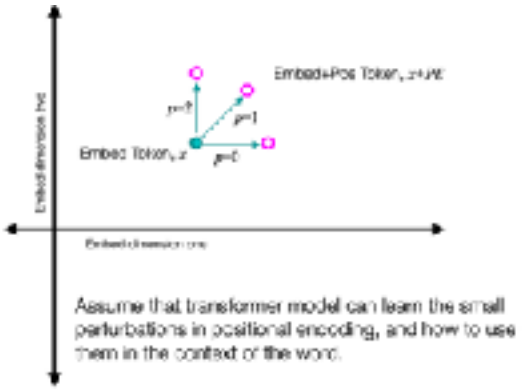
- Logistics
  - Grading update
- Agenda
  - Vision transformer and Town Hall
  - Student Paper Presentation
  - Consistency Loss
- Next Time
  - Multi-modal and Multi-Task

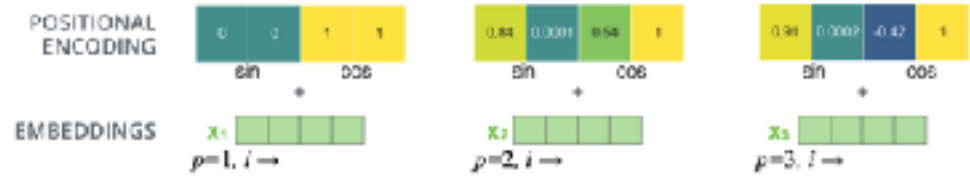
Last Time: Transformers

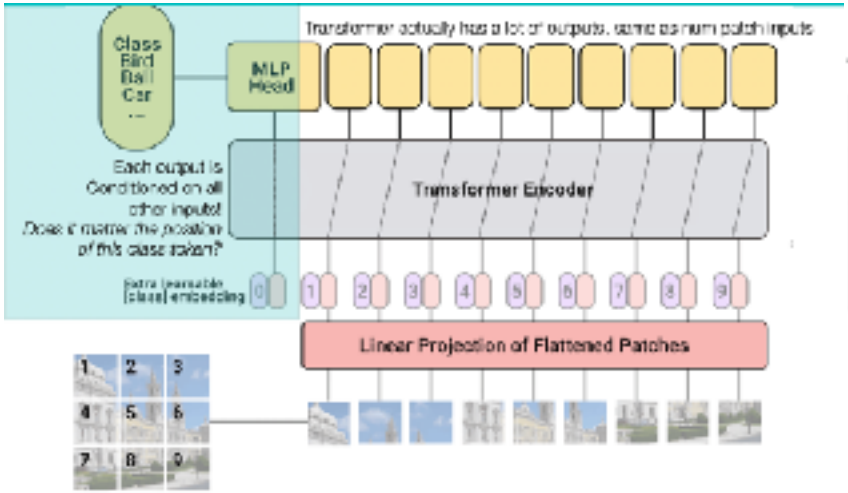
Transformer: Multi-headed Attention

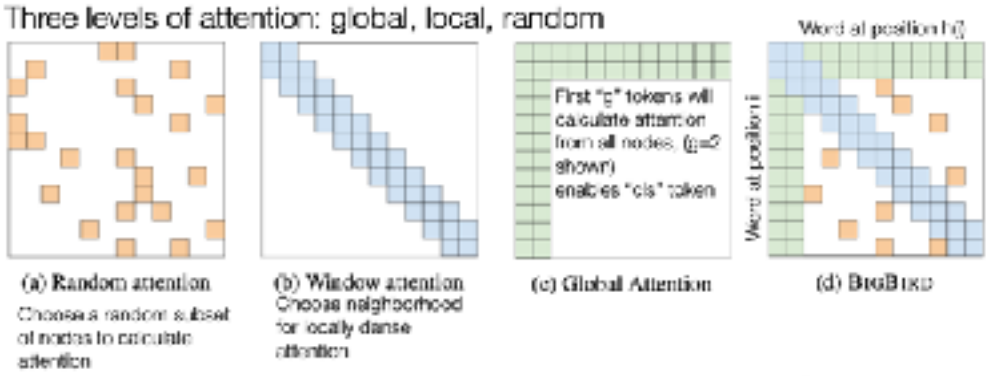




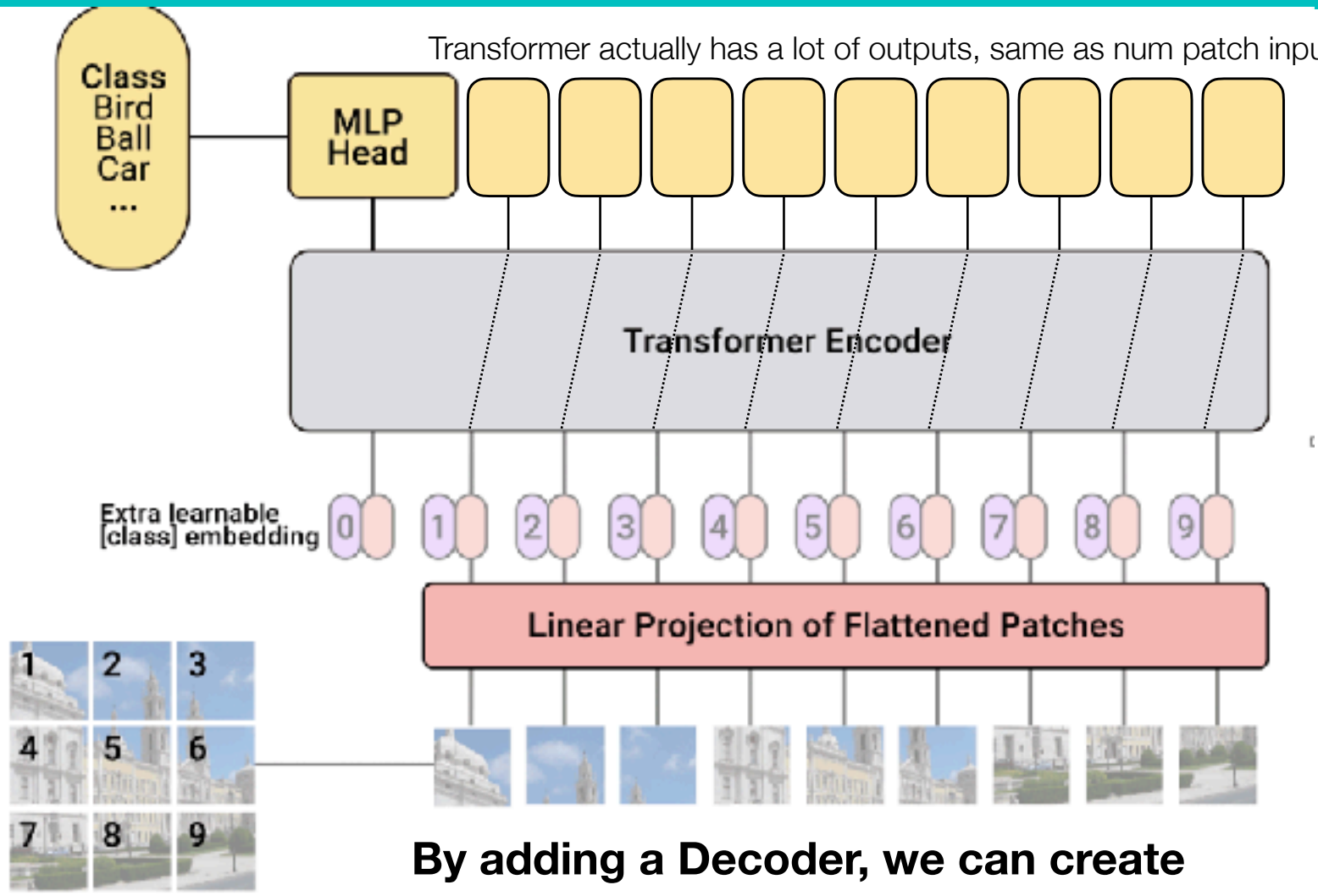








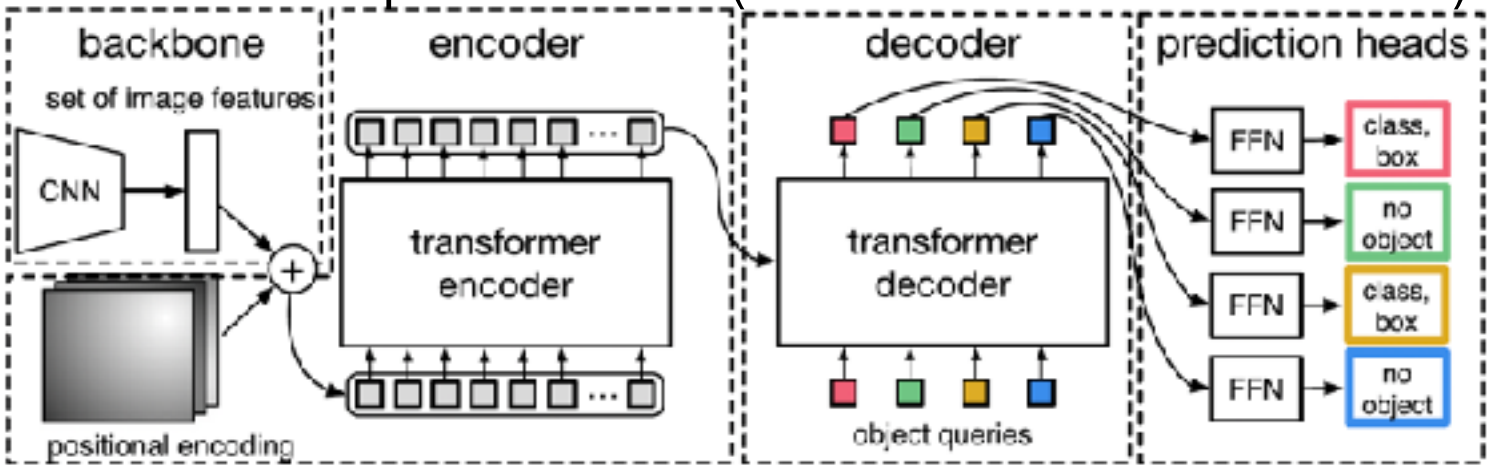
Vision Transformer Review

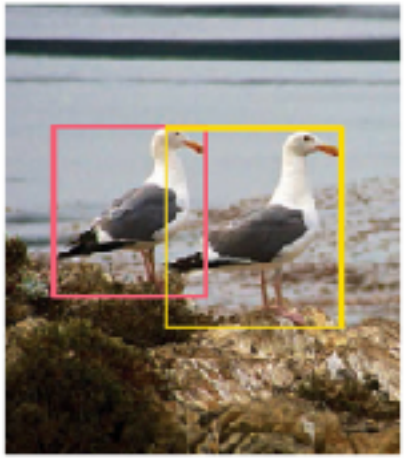


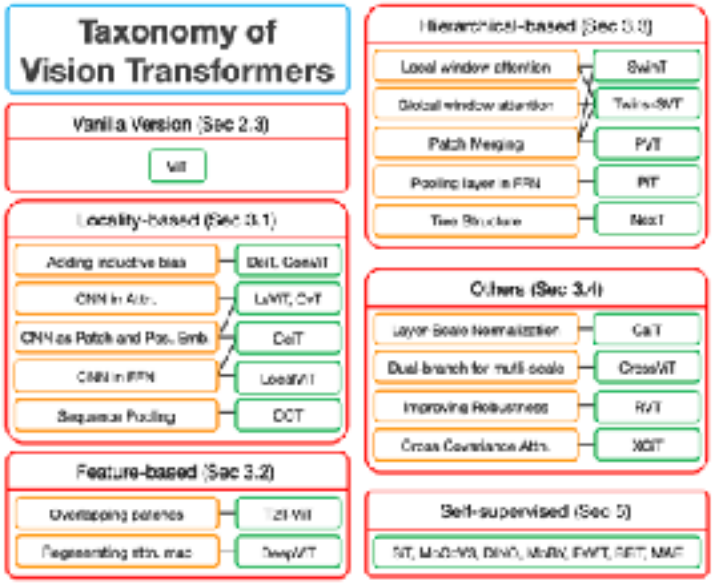
By adding a Decoder, we can create some really interesting designs!

Many Variants of the ViT

- One example: DETR (Detection Transformer)







- ViT is still an ongoing area of research
- Input Patch Structure (overlap)
- Efficient Attention, Cross Attn.
- Methods of SSL
- Image/text generation

Transformer Town Hall

Hugging Face Text transformers: https://huggingface.co/transformers/v3.3.1/pretrained_models.html
Hugging Face ViT: https://huggingface.co/docs/transformers/model_doc/vit
Keras text Transformers: https://keras.io/guides/keras_nlp/transformer_pretraining/
Keras ViT: https://github.com/faustomorales/vit-keras

Paper Presentation

VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning

Adrien Bardes[1,2] **Jean Ponce**[2,4] **Yann LeCun**[1,3,4]

[1]Facebook AI Research
[2]Inria, École normale supérieure, CNRS, PSL Research University
[3]Courant Institute, New York University
[4]Center for Data Science, New York University