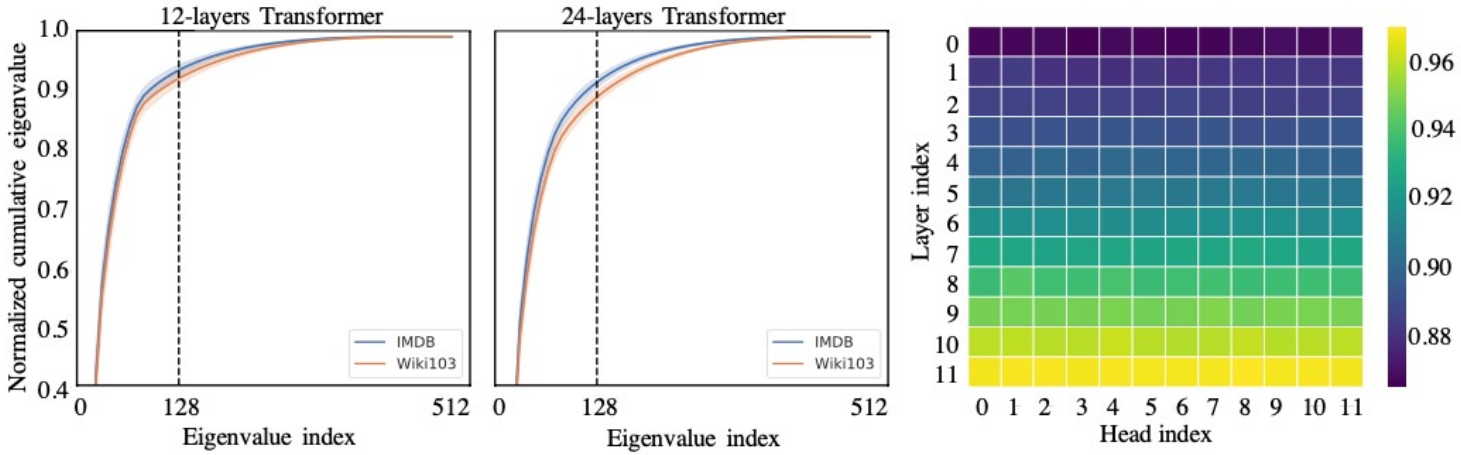
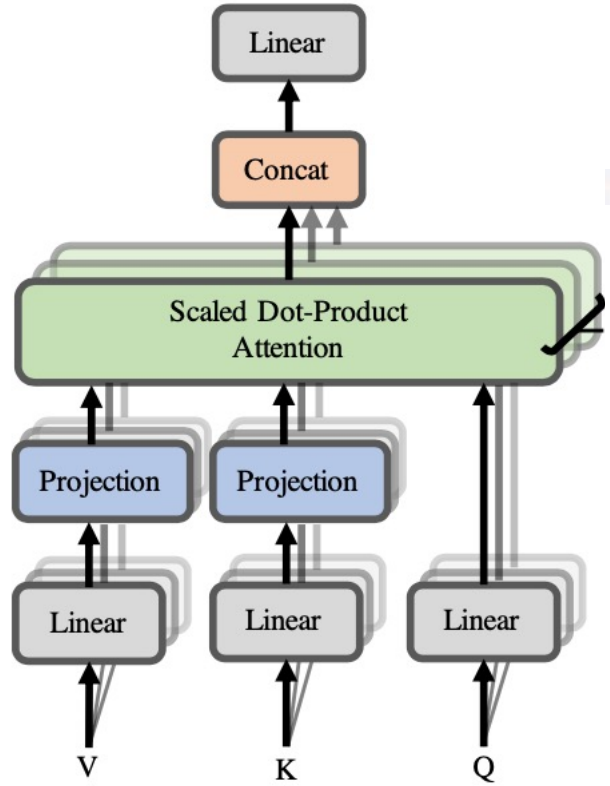
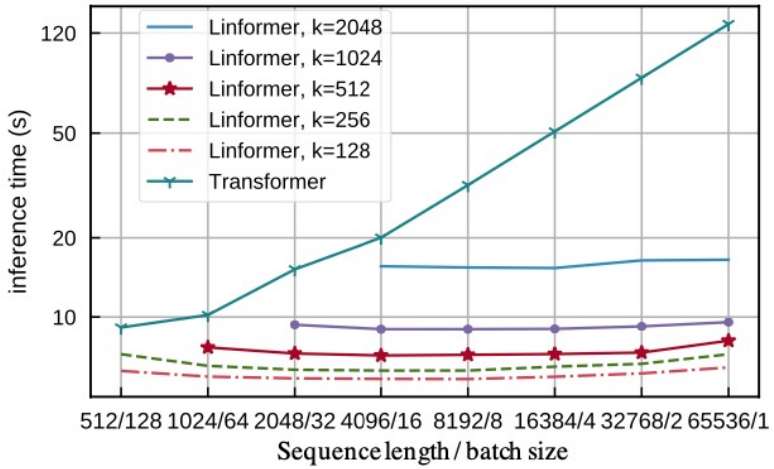
Keio University

Linformer [Wang+, 20]

- TransformerのSelf-Attention
  - 計算量O(n^2)
- KeyとValueに対しprojection layerを導入
- 特異値分解を用いて低ランク行列に近似
  - 行列の情報の殆どが
    最初の数個の最大特異値から回収可
  - 計算量O(n)





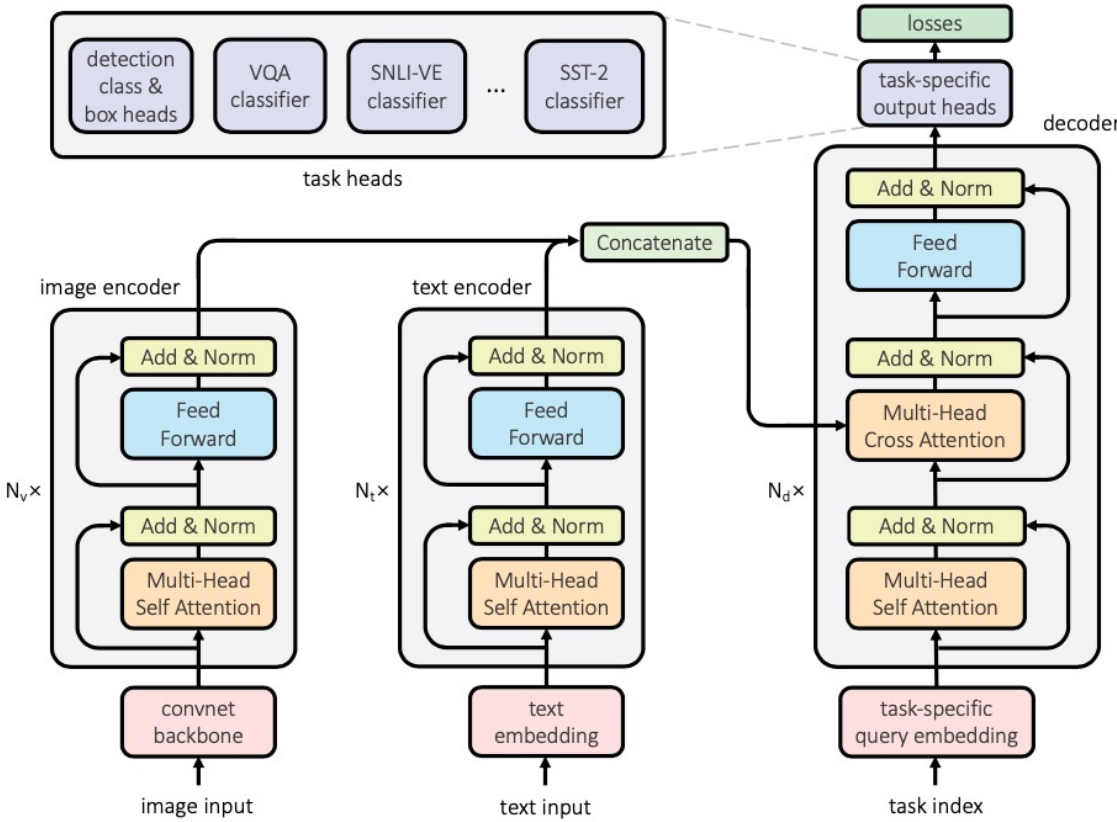


Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, Hao Ma. (2020). Linformer: Self-Attention with Linear Complexity. arXiv:2006.04768.

Keio University

UNiT [Hu(Facebook)+, 21]

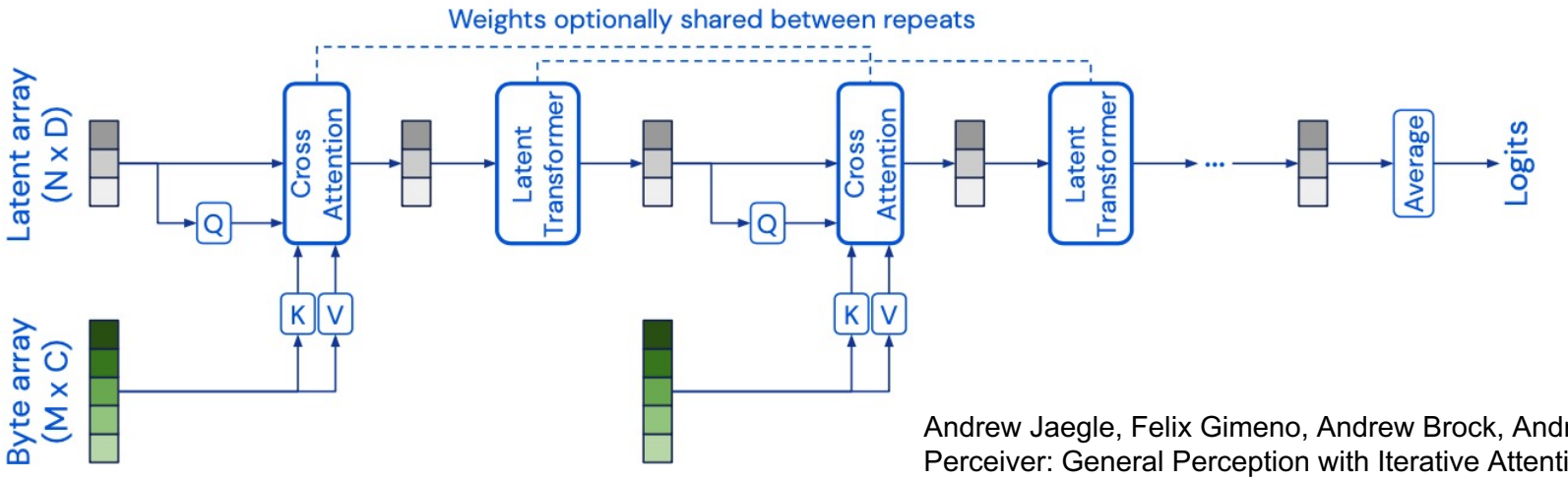
- 複数のタスクを単一モデルで学習するTransformer
  - 物体検知、自然言語理解、VQA等
- Image Encoder
  - CNNとTransformer Encoder
- Text Encoder
  - BERT
- Decoder
  - DETR
    - タスクによって必要なEncoderを使用
    - V&Lタスクでは各Encoderの出力をconcat
- Task-specific output heads
  - 各タスクがそれぞれprediction headを所持
  - すべてのタスクtをc_tクラス間の分類タスクとしてモデル化



Ronghang Hu, Amanpreet Singh. (2021). UniT: Multimodal Multitask Learning with a Unified Transformer. arXiv:2102.10772.

Perceiver [Jaegle+, 21]

- Transformerを何十万もの入力に対応
  - ImageNet：畳み込みせず50000ピクセルを処理
- Cross Attention
  - 高次元のByte array(pixel arrayなど) を低次元のLatent array(サイズはハイパーパラメータ)へ投影してから複数のTransformerで処理
- Latent Transformerの入力が低次元なため、より深いネットワーク構成が可能
  - 重みを共有し、パラメータ数を減少
  - 複数のbyte-attend層
    - Latent arrayが必要に応じて入力画像から情報を反復的に抽出可能



Andrew Jaegle, Felix Gimeno, Andrew Brock, Andrew Zisserman, Oriol Vinyals, Joao Carreira. (2021). Perceiver: General Perception with Iterative Attention. arXiv:2103.03206.