Linear Attention meets Semantic Segmentation

李睿
武汉大学遥感信息工程学院

Introduction

- Context matters for semantic segmentation
  Per-pixel classification is often ambiguous in the presence of only local information. But the task becomes much simpler if contextual information, from the whole image, is available.

- Attention is effective for context extraction
  With strong capabilities to capture long-range dependencies, dot-product attention mechanisms have been applied in vision and NLP tasks.

- Capturing context information is consuming
  Utilization of the dot-product attention mechanism often comes with significant memory and computational costs, which increases quadratically with the size of the input over space and time.

Methodology

- Dot-product attention mechanism
  \[
  D(Q, K, V) = \rho(QK^T)V.
  \]

  where
  \[
  \begin{align*}
  Q &= XW_q \in \mathbb{R}^{N \times D_k}; \\
  K &= XW_k \in \mathbb{R}^{N \times D_k}; \\
  V &= XW_v \in \mathbb{R}^{N \times D_v},
  \end{align*}
  \]

  and
  \[
  \rho(QK^T) = \text{softmax}_{\text{row}}(QK^T)
  \]

  Replace the softmax function by its Taylor expansion
  \[
  e^{q_i^T \cdot k_j} \approx 1 + q_i^T \cdot k_j
  \]

- Linear attention mechanism
  \[
  \sum_j V_{i,j} + \left( \frac{Q}{\|Q\|_2} \right) \left( \frac{K}{\|K\|_2} \right)^T V
  \]

  \[
  D(Q, K, V) = \frac{\sum_j V_{i,j} + \left( \frac{Q}{\|Q\|_2} \right) \left( \frac{K}{\|K\|_2} \right)^T V}{N + \left( \frac{Q}{\|Q\|_2} \right) \sum_j \frac{K}{\|K\|_2}_{i,j}}
  \]

Results

Fig. 2. Comparison between the (a) computational and (b) memory requirements of the linear attention mechanism and dot-product attention mechanism.

Fig 3. The mIoU of different methods on the UA Viddataset.

Conclusion

The proposed linear attention mechanism is an effective and efficient method which balances the global context and resource consumption well. Based on the linear attention mechanism, the proposed ABCNet achieves a comparative result on UA Vidd dataset.

References

