•Ù{Ç‹ºe‡0 DynamicViT: Efficient Vision Transformers with Dynamic Token Sparsification0 N-O•u(N†Nå

1. **\@•èry_•b•_q**ÿ Ql_ 1ÿ ÿ
   \[
   z_{\text{local}} = \text{MLP}(x) \in \mathbb{R}^{N \times C'}
   \]
   QvN- \(C' = C/2 \)ÿ u(NŽ–M~ô0

2. **Qh\@ry_•€ZT **ÿ Ql_ 2-3ÿ ÿ
   \[
   z_{\text{global}} = \text{Agg}(\text{MLP}(x), \hat{D}) \in \mathbb{R}^{C'}
   \]
   €ZT Qýep \(\text{Agg}\) [žs°N:^&c©x v„^sWGl`S ÿ
   \[
   \text{Agg}(u, \hat{D}) = \frac{\sum_{i=1}^N \hat{D}_i u_i}{\sum_{i=1}^N \hat{D}_i}
   \]

3. **\@•è-Qh\@ry_•‡•T N i,s‡˜„mK**ÿ Ql_ 4-5ÿ ÿ
   \[
   z_i = [z_{\text{local}_i}, z_{\text{global}}]
   \]
   \[
   \pi = \text{Softmax}(\text{MLP}(z)) \in \mathbb{R}^{N \times 2}
   \]
   •“QúkÏN*tokenˆ«OÝuYÿ \(\pi_{i,1}\)ÿ b N"_ ÿ \(\pi_{i,0}\)ÿ v„i,s‡0

4. **Gumbel-Softmax‘Çh7**ÿ Ql_ 7ÿ ÿ
   \[
   D = \text{Gumbel-Softmax}(\pi)_{*,1} \in \{0,1\}^N
   \]
   u(NŽSï_®R v„NŒP<Q³{V‘Çh70

5. **lèa R›c©x ‹¡{—**ÿ Ql_ 9-11ÿ ÿ
   \[
   P = QK^T / \sqrt{C}
   \]
   \[
   G_{ij} = \begin{cases}
   1 & \text{if } i=j \\
   \hat{D}_j & \text{if } i \neq j
   \end{cases}
   \]
   \[
   \tilde{A}_{ij} = \frac{\exp(P_{ij})G_{ij}}{\sum_{k=1}^N \exp(P_{ik})G_{ik}}
   \]
   • •Çc©x wé–5 \(G\) –;e-ˆ«N"_ tokenv„N¤N’0

6. **‹-~Ãvîh **ÿ Ql_ 12-16ÿ ÿ
   - R |{c_Y1ÿ \(L_{\text{cls}} = \text{CrossEntropy}(y, \bar{y})\)
   - •ê„¸™•c_Y1ÿ \(L_{\text{distill}} = \frac{1}{\sum \hat{D}_i} \sum \hat{D}_i (t_i - t'_i)^2\)
   - KLec^¦c_Y1ÿ \(L_{\text{KL}} = \text{KL}(y \| y')\)
   - kÔO‹~¦g_c_Y1ÿ \(L_{\text{ratio}} = \frac{1}{BS} \sum_{b,s} (\rho^{(s)} - \frac{1}{N} \sum_i \hat{D}_i^{b,s})^2\)
   `;c_Y1N:R gCTŒÿ \(L = L_{\text{cls}} + \lambda_{\text{KL}} L_{\text{KL}} + \lambda_{\text{distill}} L_{\text{distill}} + \lambda_{\text{ratio}} L_{\text{ratio}}\)0

•ÙN›Ql_ QqT [žs°N†R¨` tokenz u•S v„h8_Ãg:R6ÿ S bì‘Í‰•`˜„mK0 Sï_®R Rjg•TŒxlNöSËY}v„lèa R›c©x