•Ù{Ç‹ºe‡0 DynamicViT: Efficient Vision Transformers with Dynamic Token Sparsification0 cÐQúN†N yÍ
Transformerÿ ViTÿ N-šØeHW0Rjg•Q—OYNärL0 NåN f/‹ºe‡v„N;‰•Q…[¹TŒ•!s.ÿ

1. h8_Ã` `ó
- **‰Â[ß**ÿ ‰Æ‰ÉTransformerv„g ~È˜„mKNÅO••VNŽN \ •èR Oá`o‘Ïg Y'v„NärLÿ tokensÿ ÿ QvNÖQ—
- **e¹lÕ**ÿ cÐQúR¨` NärLz u•S hFg¶ÿ • •Ç•{‘Ï~§˜„mKj!WWÿ prediction moduleÿ • \BRjg•Q—OYNärLÿ N

2. Qs•.b€g/
- **R¨` NärLRjg•**ÿ
  - ˜„mKj!WWWúNŽ_SRMry_•O0‹¡kÏN*NärLv„‘Í‰•`'R epÿ u b NŒ•ÛR6Q³{Vc©x ÿ binary decision mask
  - NärLRjg•f/\Bk!S v„ÿ hierarchicalÿ ÿ • keXžR Rjg•kÔO‹0
- **zïR0zïO S **ÿ
  - O•u(Gumbel-Softmax‰ãQ³—^Sï_®v„‘Çh7•î˜˜0
  - cÐQúlèa R›c©x {Vueÿ attention maskingÿ ÿ • •Ç–;e-ˆ«Rjg•NärLN QvNÖNärLv„N¤N’ÿ [žs°Sï^vˆL‹¡{—
- **‹-~Ãvîh **ÿ
  - ~ÓT R |{c_Y1ÿ cross-entropyÿ 0 „¸™•c_Y1ÿ self-distillationÿ 0 KLec^¦c_Y1ÿ KL divergenceÿ TŒkÔ

3. [žšŒ~Ógœ
- **`'€ý**ÿ
  - W(ImageNetN ÿ Rjg•66%v„NärLT ÿ FLOPsQÏ\ 31%~37%ÿ T T ‘ÏcÐSG40%NåN ÿ |¾^¦N –MN •…•Ç
  - W(DeiT0 LV-ViT{Ij!W‹N šŒ‹ÁN†g eH`'0
- **O R¿**ÿ
  - vøkÔ~Óg„S Rjg•ÿ Y,CNNv„l`S ÿ ÿ R¨` Rjg•fôpum;N xlNöSËY}0
  - Sï‰ÆS f>y:j!W‹€ý• ke€Zq&NŽVþPÏN-v„Qs•.S:Wß0

4. R e°p¹
- **R¨` `'**ÿ Rjg•Q³{VWúNŽ•“QeQ…[¹ÿ € —^Vú[šj!_ 0
- **šØeH`'**ÿ • •Çlèa R›c©x [žs°xlNöSËY}v„z u•‹¡{—0
- **• u(`'**ÿ Sï^”u(NŽY yÍ‰Æ‰ÉTransformerg¶g„ÿ Y,ViT0 DeiT0 LV-ViTÿ 0

5. ^”u(N \Ug
- • u(NŽVþPÏR |{NûR¡ÿ g*geSïbi\U•ó‰Æ˜‘R |{TŒ[Æ–Æ˜„mKNûR¡0
- Nãx _ n•ÿ https://github.com/raoyongming/DynamicViT0

`;~Ó
DynamicViT• •ÇR¨` Rjg•Q—OYNärLÿ f>„WcÐSGN†‰Æ‰ÉTransformerv„eHs‡ÿ T eöOÝc N†šØ|¾^¦ÿ