

### 0 DynamicViT: Efficient Vision Transformers with Dynamic Token Sparsification0 %ã'ûb¥TJ`;-Ó

#### 1. \*\*x zv€lfoN h8\_Ã•ĩ~\*\*

- \*\*€lfo\*\*y Vision Transformersy ViTy W(VpPİR |{{INûRjN-^hs°O \_ y OFQv{q—Y gB^!šØy \ \$Qvf/•êlèa I
- \*\*h8\_Ã•ĩ~\*\*y Y,OUW(N f>„W—MONj!W'€ýv„RMcDN y šØeHQİ\ ViTv„,q{—İy

#### 2. \*\*Qs•.%oÂ[ßN R e°p1\*\*

- \*\*Qs•.%oÂ[ß\*\*y ViTv„g ~È~„mKNÂO••VNŽ\ 'İOá`o'İšØv„VpPİWWy tokeny y QvOYtokenQ—OY^!šØ0
- \*\*R e°p1\*\*y
  - \*\*R`` tokenz u•S \*\*y cĐQúWúNŽ“QeQ...[!v„R`` tokenRjg•hFg¶İy DynamicViTy y • keRT—dQ—OYto
  - \*\*•{!~„mKj!WW\*\*y ••ÇMLP~„mKkİN\*tokenv„!%o•`R epy ~ÓT \@•èN Qh\@ry\_•u b Rjg•Q³V0
  - \*\*lèa R>c@x {Vue\*\*y ••ÇSİ\_@v„Gumbel-SoftmaxTœlèa R>c@x y \O...=^«Rjg•tokenv„N¤N'y y [žs°ziF

#### 3. \*\*e!İÖ~Æ„\*\*

- \*\*R \BRjg•\*\*y W(Y N\*Transformer\BcÔQe~„mKj!WWy R —6kµRjg•y Y,12\Bj!W'N-y W({,40 70 10\BRjg
- \*\*(-~Ã{Vue\*\*y
  - \*\*c\_Y1Qýep\*\*y €TT O S R |{c\_Y1y CrossEntropyy 0 „,™•c\_Y1y N eY^ j!W'token[ùYPy 0 KLec^!y •
  - \*\*c"t R • \*\*y vœçc ~„,¾kÔO'ÖYUšØ'Í%o•`tokeny eà— € †QSi\_@'0

#### 4. \*\*[žšœ~Ógœ\*\*

- \*\*`'€ý[ùkÔ\*\*y
  - \*\*R • eHgœ\*\*y Rjg•66%v„tokeny OYUy34%y y FLOPsQİ\ 31%~37%y T T İcĐSG40%+y |¾^!N —M<
  - \*\*j!W'ùkÔ\*\*y DynamicViT-LV-M/0.7W(8.5G FLOPsN •¾R083.8%QÆxns†y O NžEfficientNet-B5y 9
- \*\*Si%oÆS R g•\*\*y Rjg•T j!W'€Zq&NžVpPİQs•.S:Wßy Y,riOSN-\_Äy y šœ'ÂN†R`` Rjg•v„T t `0

#### 5. \*\*b€g/O Rç\*\*

- \*\*xİNöSËY}\*\*y —^~Óg„S Rjg•y OYUytokenOM•nN Vú[šy NÍSi• •Ç•êlèa R>g:R6^v^L{q{—0
- \*\*pum;`'\*\*y e/c N T ViTSØOSy Y,DeiT0 LV-ViTy y N • u(NžY'j!W'y Y,384×384•“Qey 0

#### 6. \*\*\@—PN g\*gee1T \*\*

- \*\*NûRjbi\U\*\*y \_SRM€Zq&R |{NûRjy g\*geSicø}"%oÆ~'R |{0 [Æ—Æ~„mKy Y,hÂmK/R Rry 0
- \*\*t ¸°R g•\*\*y Rjg•[ùj!W'œ•hÒ`TœSi%ã'È`v„\_qTÍ— •ÛN kex zv0

#### 7. \*\*`;~Ó\*\*

DynamicViT••ÇR`` tokenz u•S y ™—k![žs°N†ViTv„šØeHR • N `€ý^s^ay N:TransformerW(%oÆ%oÉNûR

y lèy Y,— •ÛN keR g•QwOS[žšœ~Æ„,b biU^"u(W:foý Siœ u(vøQs]âQwf•SÖ^eQEepcn0 y