Designing Network Design Spaces

Thursday, 9. September 2021 3:57 PM

- Adresses NAS
- Creates RegNet architecture
- Width and depth of good nets can be explained by a quantized linear function
- Allows for good tradeoff between accuracy and speed (5x more efficient than EfficientNet)
- Used by Tesla AP team as vision backbone
- Combines NAS and manual design (NAS only omptimizes a single architecture)
- Progressively design choices (RegNet design space) based on an initaly very unconstraint design space (AnyNet) , then explore width, depth, blocks
- Efficient and good for mobile applications
- Start with unconstraint design, then introduce priors
- Compared in low ressource low epoch regime for efficiency
- Error used :

As in [21], our primary tool for analyzing design space quality is the error *empirical distribution function* (EDF). The error EDF of n models with errors e_i is given by:

$$F(e) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}[e_i < e]. \qquad (1)$$

$F(e)$ gives the fraction of models with error less than e. We

- Steps :

1) generate and train several models
2) Compute and plot EDF
3) Visualize properties use empirical bootstrap
4) Refine design space

- AnyNetx design space : 16 DoF (block width, group with, bottleneck ratio, number of blocks)
- Do not compute all configs (10^18), but rather explore design principles
- Observations : stage widthss/depths increase for models
- RegNet design space :

=> linear fit of models in AnyNetXe emerges

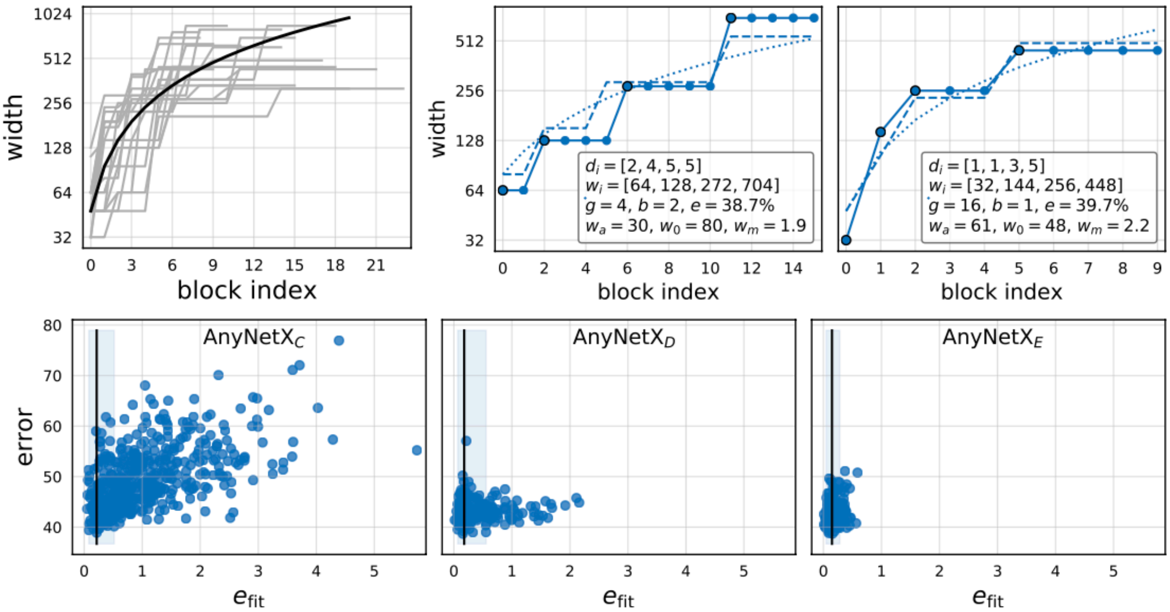

Figure 8. **Linear fits.** Top networks from the AnyNetX design space can be well modeled by a *quantized linear parameterization*, and conversely, networks for which this parameterization has a higher fitting error e_{fit} tend to perform poorly. See text for details.

Propose :

1) linear parameterization of block width
2) Param to control quantization (block width control)
3) Minimize mean log ration of per block widths

	restriction	dim.	combinations	total
AnyNetX$_A$	none	16	$(16{\cdot}128{\cdot}3{\cdot}6)^4$	$\sim 1.8{\cdot}10^{18}$
AnyNetX$_B$	$+\, b_{i+1} = b_i$	13	$(16{\cdot}128{\cdot}6)^4{\cdot}3$	$\sim 6.8{\cdot}10^{16}$
AnyNetX$_C$	$+\, g_{i+1} = g_i$	10	$(16{\cdot}128)^4{\cdot}3{\cdot}6$	$\sim 3.2{\cdot}10^{14}$
AnyNetX$_D$	$+\, w_{i+1} \geq w_i$	10	$(16{\cdot}128)^4{\cdot}3{\cdot}6/(4!)$	$\sim 1.3{\cdot}10^{13}$
AnyNetX$_E$	$+\, d_{i+1} \geq d_i$	10	$(16{\cdot}128)^4{\cdot}3{\cdot}6/(4!)^2$	$\sim 5.5{\cdot}10^{11}$
RegNet	quantized linear	6	$\sim 64^4{\cdot}6{\cdot}3$	$\sim 3.0{\cdot}10^{8}$

Table 1. **Design space summary.** See text for details.

Findings :

1) Optimal depth of 20 blocks
2) Bottleneck ratio of 1.0
3) Width multiplier is 2.5
4) Use of inverted bottleneck is worse

Depthwise Conv and Swish interact favorably