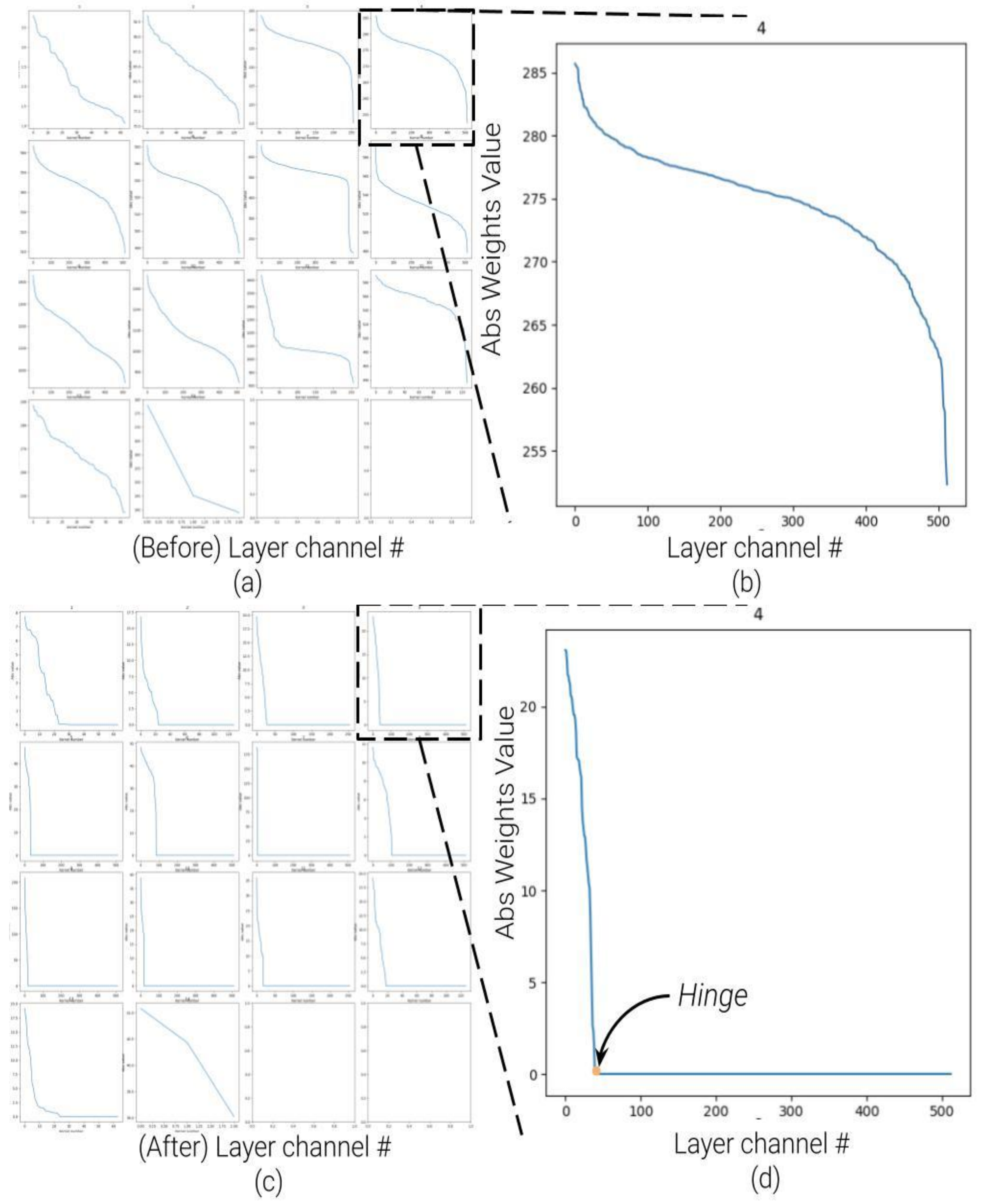
Towards Device Efficient Conditional Image Generation

Nisarg A. Shah and Gaurav Bharaj
AI Foundation, California, USA

AI FOUNDATION

BMVC 2022

Motivation

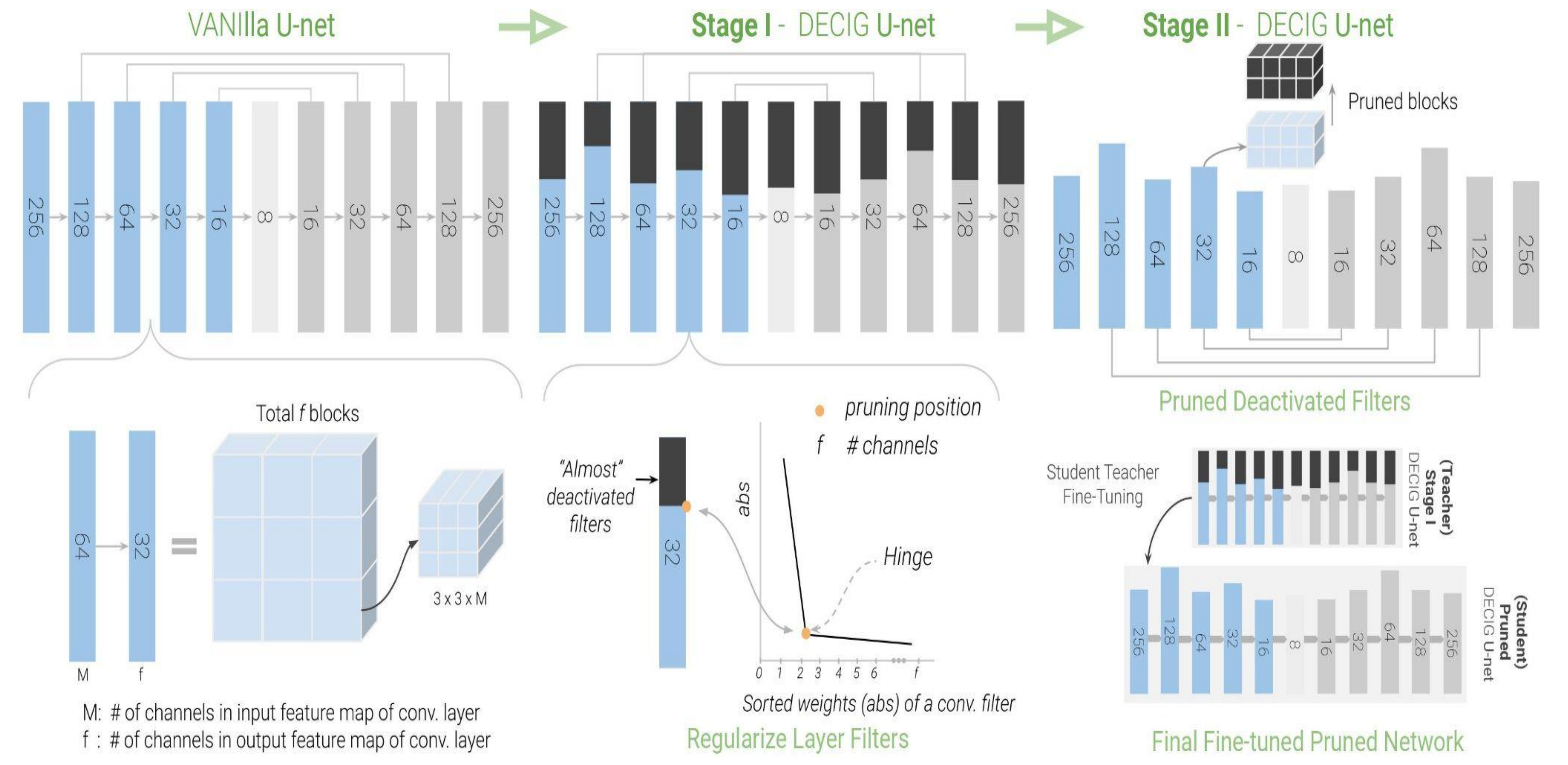


Goal : Efficient and accurate representation of model weights, for superior channel pruning

Our Approach

- We propose a two-stage novel strategy where, first, we condense the channel weights, such that, as few channels are used.
- Later we prune, nearly zeroed out weight activations, and fine-tune the autoencoder.
- To maintain image quality, fine-tuning is done via student-teacher training, condensed model - (Teacher)

Network Architecture



Method Overview

- We propose, channel weight and layer device performance device regularisation, both operating at intra and inter-layer level
- Channel importance factor γ, is equivalent to magnitude of the weights of the corresponding channels.
- We calculate the run-time for each layer across a particular device, and use it as a multiplicative factor l(i) for that layer. (device-dependent)

'Hinge' based pruning

- On Stage-I training, a model with a considerable amount of near zero weight channels are obtained with considerable distinction.
- The inclination point that shows the threshold between these two types of channels is identified as the ``hinge".
- In turn, not requiring to take an arbitrary guess or a global threshold on the number of channels to be pruned.

Channel level : $L_i = \Sigma_{j=1}^{n} f(j) * ||W_{i,j}||_1$

Layer level : $L_{PENAL} = \Sigma_{i=0}^{n} l(i) * L_i$

$W_{i,j}$ = Filter weight of i^{th} layer and j^{th} sorted channel
$f(j)$ = channel regularsation function; Linear, Uniform, ..
$l(i)$ = Runtime for layer on particular device

Experiments and Results

- Generator models such that UNet and ResUNet, and there corresponding DECIG versions were used
- Inference times are calculated for both CPU and GPU
- Conditional image generation tasks – Segmentation mask to images, images to cartoonization and CycleGAN

Segmentation Mask	Ground Truth	Vanilla Unet-16	Vanilla Unet-32	Vanilla Unet-64	DECIG Unet-64 *High reg.*	DECIG Unet-64 *Low reg.*
CPU (FPS) ↑		**70**	25.9	7.3	25.2	16.4
GPU (FPS) ↑		**200**	168	25.9	156	131
FID ↓		74.5	58.7	47.3	48.6	**37.9**
Params (M) ↓		**2.62**	10.46	41.83	3.74	9.5

Source Image	UNet-64	ResNet	Shu et al.	Wang et al.	DECIG-ResNet
FID ↓	138	71.87	96	88	**62.72**
Flops (Giga) ↓	**6.03 G**	52.90 G	12.51 G	11.36 G	10.9 G
Memory (MB) ↓	160 MB	43.51 MB	10.19 MB	8.7 MB	**7.42 MB**

Face Images	Ground Truth	UNet-64	DECIG-UNet-64 Linear (low reg.)	DECIG-UNet-64 Uniform(Stage-1)	DECIG-UNet-64 Uniform(Stage-2)
FID ↓		29.20	27.52	25.65	**24.31**
Parameters (Million) ↓		41 M	18.5 M	41 M	**7.8 M**
Memory (MB) ↓		160 MB	96 MB	160 MB	**30 MB**
CPU (FPS) ↑		7.4	8.2	7.4	**11.6**
GPU (FPS) ↑		96	107	96	**119**