BMVC 2023

aselsan

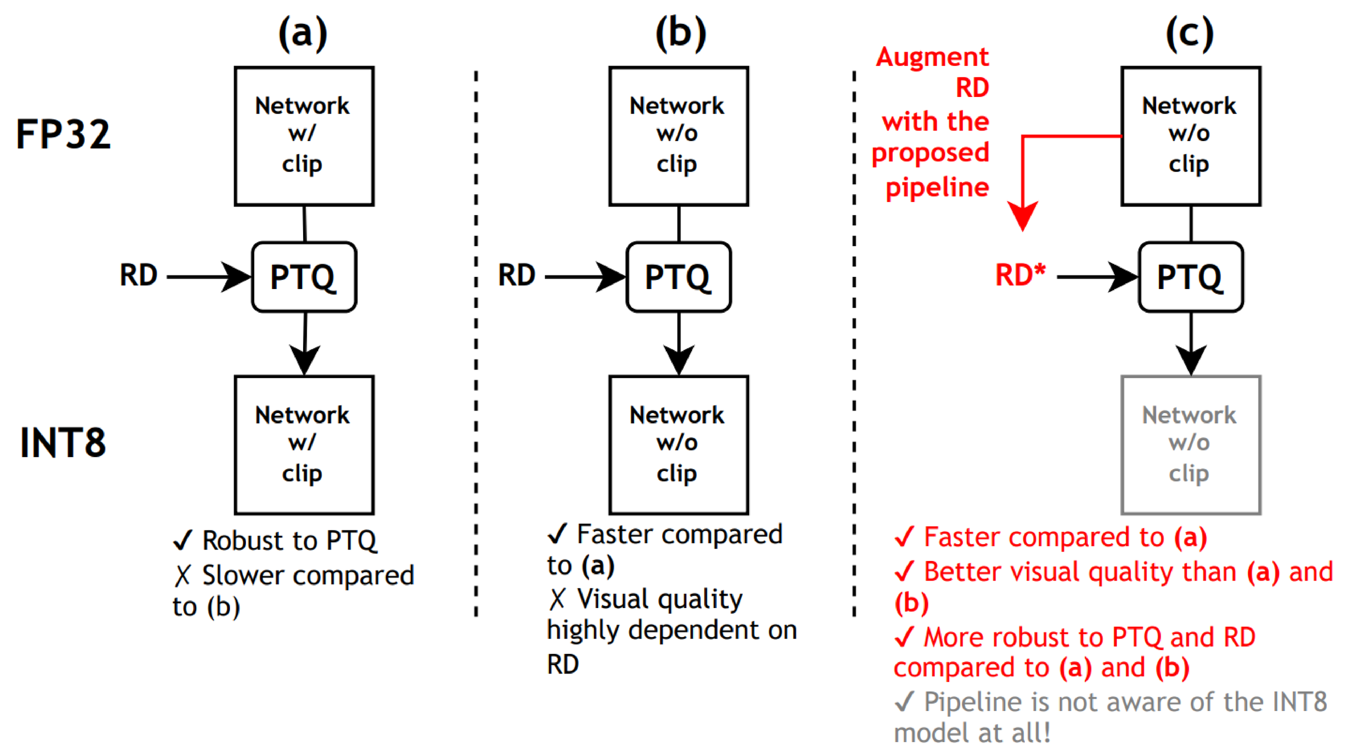
Towards Clip-Free Quantized Super-Resolution Networks: How to Tame Representative Images

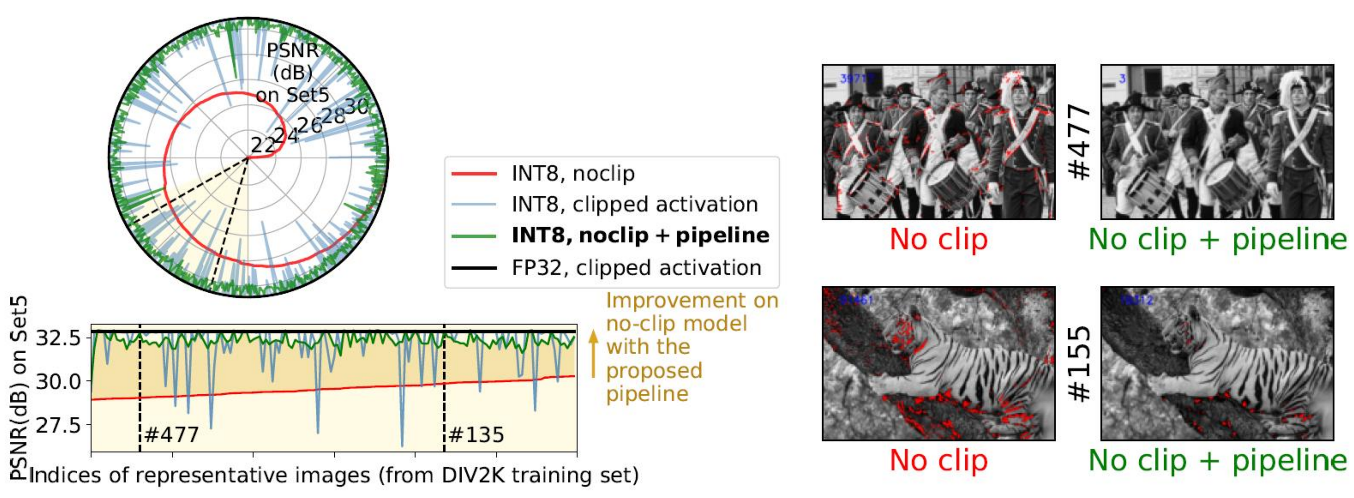
Alperen Kalay, Bahri Batuhan Bilecen, Mustafa Ayazoglu
Aselsan Inc., Türkiye

Introduction

Nearly all quantized mobile super-resolution (SR) networks utilize clipped activations to make the model more robust to post-training quantization (PTQ) in return for a large overhead in runtime. We propose a novel pipeline to **augment representative dataset (RD) used in PTQ, which can successfully eliminate unwanted clipped activation layers.**

Removing clipped activations with our method significantly benefits **decreased inference runtime up to 54% on some SR models**, better visual quality results compared to INT8 clipped models - and **outperforms even some FP32 non-quantized models**, both in runtime and visual quality, without the need for retraining with clipped activation.

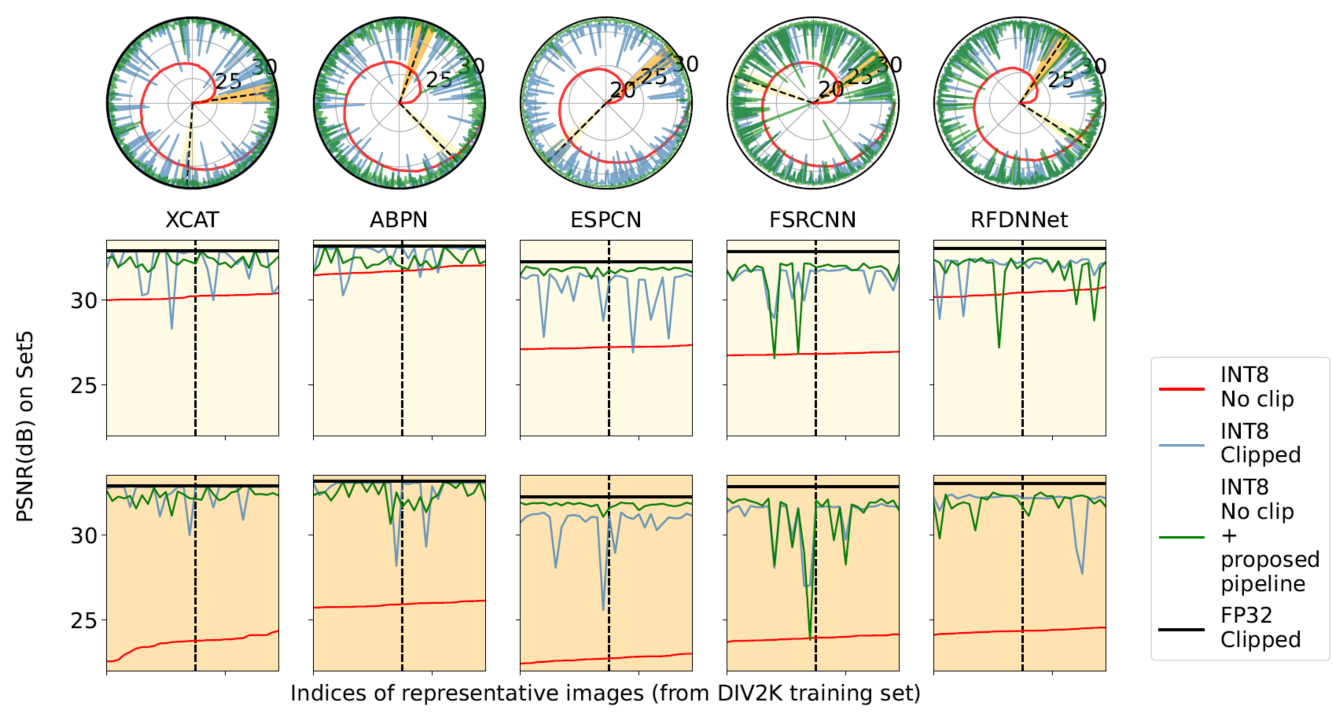




Qualitative & Qualitative Results

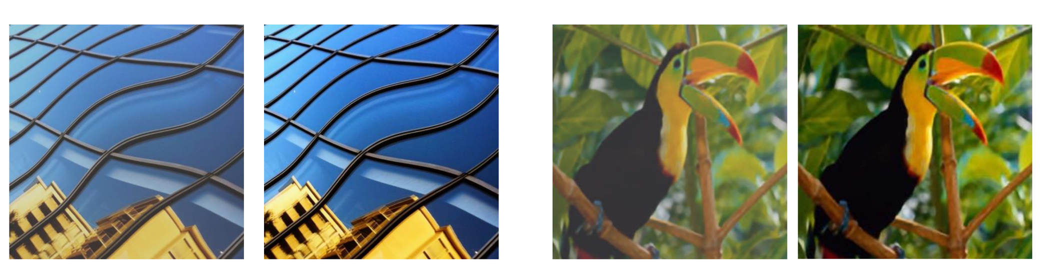
	XCAT ✓	XCAT ✗	XCAT ↑	ABPN ✓	ABPN ✗	ABPN ↑	ESPCN ✓	ESPCN ✗	ESPCN ↑	FSRCNN ✓	FSRCNN ✗	FSRCNN ↑	RFDNNet ✓	RFDNNet ✗	RFDNNet ↑
CPU	458	379	20.8%	779	715	9.0%	328	213	**54.0%**	1610	1564	2.9%	904	810	11.6%
TFLite GPU	460	445	3.4%	790	781	1.2%	342	339	0.9%	667	554	20.4%	484	465	4.1%
Android NNAPI	451	416	8.4%	254	197	28.9%	187	155	20.6%	309	275	12.4%	600	545	10.1%
Qualcomm Hexagon	377	338	11.5%	178	150	18.7%	113	91	24.2%	1207	1206	0%	519	498	4.2%
Qualcomm QNN GPU	502	375	33.9%	772	728	6.0%	322	213	51.2%	1605	1552	3.4%	917	836	9.7%

Runtime improvements when clipped activations are removed.



Test	Models	Clipped M1	Clipped ~~M1~~	No-Clip M1	No-Clip ~~M1~~	No-Clip ~~M1~~ → M2
Set5	XCAT	32.329 *(15)*	32.310 *(785)*	32.640 *(15)*	28.409 *(785)*	**32.346** *(785)*
	ABPN	32.492 *(15)*	32.526 *(785)*	32.975 *(15)*	28.360 *(785)*	**32.276** *(785)*
	ESPCN	31.139 *(8)*	30.855 *(792)*	31.501 *(8)*	26.213 *(792)*	**31.758** *(792)*
	FSRCNN	31.232 *(13)*	31.124 *(787)*	31.529 *(13)*	27.178 *(787)*	**31.439** *(787)*
	RFDNNet	32.167 *(7)*	31.766 *(793)*	32.079 *(7)*	27.012 *(793)*	**31.798** *(793)*
Set14	XCAT	28.956 *(15)*	28.890 *(785)*	29.131 *(15)*	26.516 *(785)*	**28.910** *(785)*
	ABPN	29.042 *(15)*	28.956 *(785)*	29.360 *(15)*	26.380 *(785)*	**28.906** *(785)*
	ESPCN	28.239 *(8)*	28.616 *(792)*	28.478 *(8)*	25.089 *(792)*	**28.649** *(792)*
	FSRCNN	28.096 *(13)*	28.149 *(787)*	28.241 *(13)*	25.539 *(787)*	**28.271** *(787)*
	RFDNNet	29.058 *(7)*	28.906 *(793)*	28.877 *(7)*	25.605 *(793)*	**28.748** *(793)*
BSD100	XCAT	28.094 *(15)*	28.126 *(785)*	28.216 *(15)*	25.907 *(785)*	**28.126** *(785)*
	ABPN	28.165 *(15)*	28.196 *(785)*	28.408 *(15)*	26.069 *(785)*	**28.150** *(775)*
	ESPCN	27.494 *(8)*	27.821 *(792)*	27.684 *(8)*	24.960 *(792)*	**27.843** *(792)*
	FSRCNN	27.611 *(13)*	27.736 *(787)*	27.723 *(13)*	25.452 *(787)*	**27.785** *(787)*
	RFDNNet	28.231 *(7)*	28.149 *(793)*	28.165 *(7)*	25.459 *(793)*	**28.073** *(793)*
Urban100	XCAT	25.964 *(15)*	25.967 *(785)*	26.137 *(15)*	24.328 *(785)*	**25.845** *(775)*
	ABPN	26.197 *(15)*	26.261 *(785)*	26.440 *(15)*	24.318 *(785)*	**25.943** *(775)*
	ESPCN	25.211 *(8)*	25.104 *(792)*	25.349 *(8)*	23.056 *(792)*	**25.448** *(792)*
	FSRCNN	25.410 *(13)*	25.444 *(787)*	25.420 *(13)*	23.537 *(787)*	**25.397** *(787)*
	RFDNNet	26.046 *(7)*	25.922 *(793)*	25.971 *(7)*	23.619 *(793)*	**25.815** *(793)*

Average PSNR improvements of the two-stage (M1, M2) pipeline.



24.64/0.914 33.00/0.948 21.21/0.896 33.40/0.958

PSNR/SSIM for sample images on INT8 no-clip vs INT8 no-clip + pipeline models.

Preliminary Observations

We have run extensive experiments to understand the underlying problem:

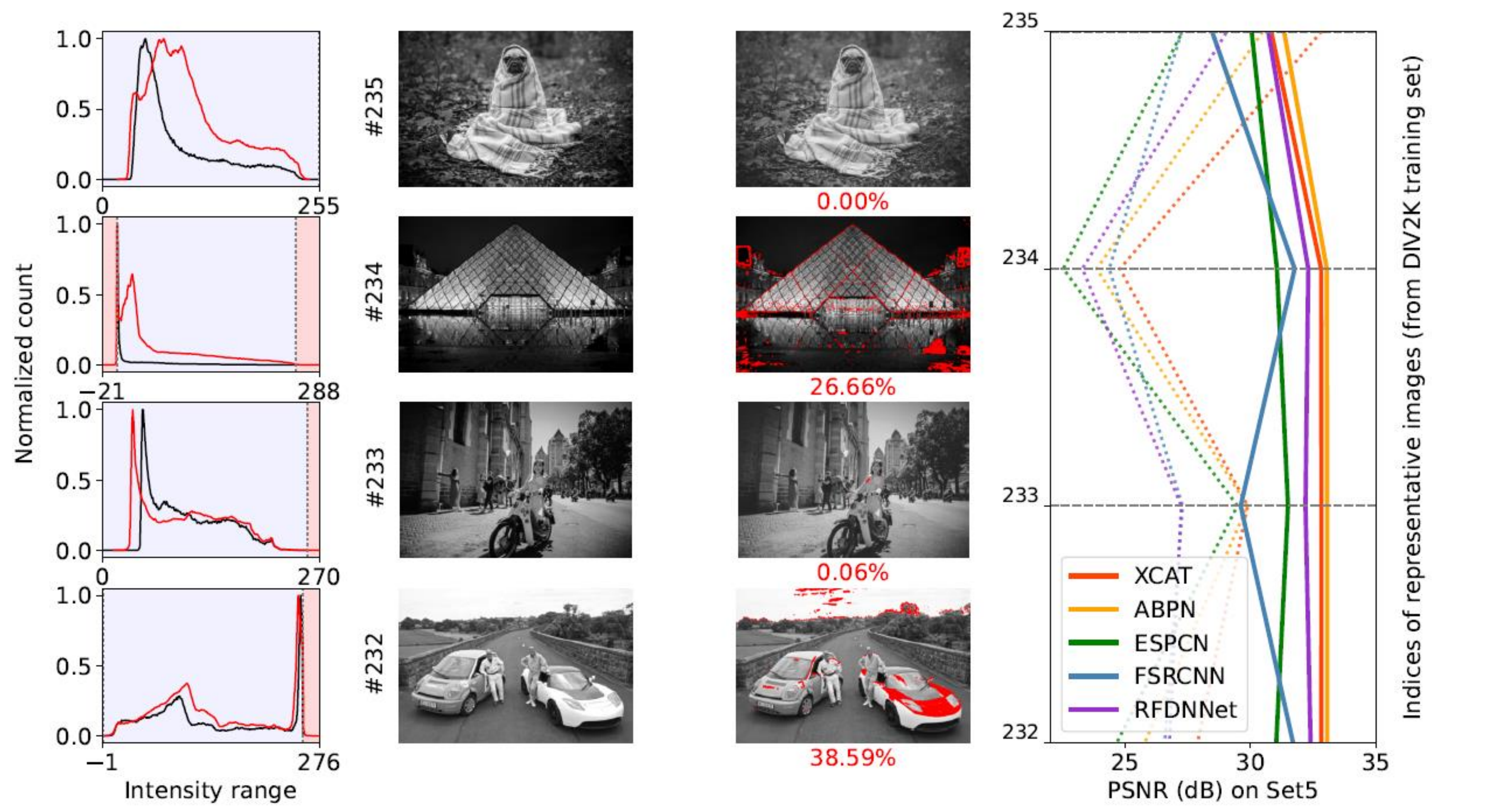
- Clipped activations affect PTQ stability drastically. Using grayscale images as the RD works better than RGB:

Models	XCAT				ABPN				ESPCN				FSRCNN				RFDNNet			
Precision	FP32		INT8		FP32		INT8		FP32		INT8		FP32		INT8		FP32		INT8	
Clip	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
$PSNR_{R=RGB}$	32.87	32.98	32.63	26.57	33.13	33.18	32.85	26.13	32.23	32.23	31.18	24.92	32.83	32.83	31.35	25.69	33.02	32.97	32.23	25.62
$PSNR_{R=Gray}$			32.31	28.48			32.52	28.44			30.85	26.26			31.12	27.24			31.76	27.05
$\Delta_{R=RGB}$	-0.11		6.06		-0.05		6.72		~0		6.26		~0		5.66		-0.05		6.61	
$\Delta_{R=Gray}$			**3.83**				**4.08**				**4.59**				**3.88**				**4.71**	

- The number of images in the RD also affects the PTQ quality. Using a single image is better since a single *bad* one can affect the overall dataset *(good RD images are the ones that have the smallest PSNR difference between their clipped INT8 and no-clip INT8 models, and vice versa for the bad ones)*

RD Size	RD	XCAT all RD	min	max	avg	ABPN all RD	min	max	avg	ESPCN all RD	min	max	avg	FSRCNN all RD	min	max	avg	RFDNNet all RD	min	max	avg
5	*Good* images	**32.84**	32.57	32.85	32.72	**33.11**	32.86	33.11	33.00	**31.44**	30.86	31.82	31.53	**32.03**	30.41	32.01	31.40	**32.28**	29.79	32.16	31.54
5	*Bad* images	**22.95**	22.95	25.70	24.05	**20.93**	20.93	24.84	23.41	**20.20**	20.20	23.43	21.94	**21.48**	21.47	24.85	23.49	**21.60**	21.72	24.09	23.16
5	Random selections from *good* and *bad*	**23.69**	23.47	32.85	29.50	**24.22**	24.39	33.11	29.67	**21.74**	21.83	31.79	28.04	**22.40**	21.83	31.79	28.04	**23.08**	21.83	31.79	28.04
		22.95	22.95	31.71	27.22	**20.93**	20.93	33.04	26.87	**20.20**	20.20	31.82	25.46	**21.47**	21.47	31.80	26.60	**21.71**	21.72	31.89	26.11
		22.95	22.95	32.86	28.95	**20.93**	20.93	33.04	28.83	**20.20**	20.20	31.79	27.21	**21.47**	21.47	31.49	27.77	**21.72**	21.72	32.04	27.89
10	*Good + bad*	**22.95**	22.95	32.85	28.38	**20.93**	20.93	33.11	28.20	**20.20**	20.20	31.82	26.73	**21.48**	21.47	32.02	27.44	**21.60**	21.72	32.16	27.35
100	DIV2K training images	**22.08**	22.08	32.85	28.19	**21.61**	21.61	33.06	28.13	**19.87**	19.87	31.79	26.04	**19.40**	31.96	19.39	26.89	**20.75**	20.72	32.04	26.83

- **The number of outliers in the FP32 model's response to the corresponding representative image gives us a measure of whether the image is *good or bad***



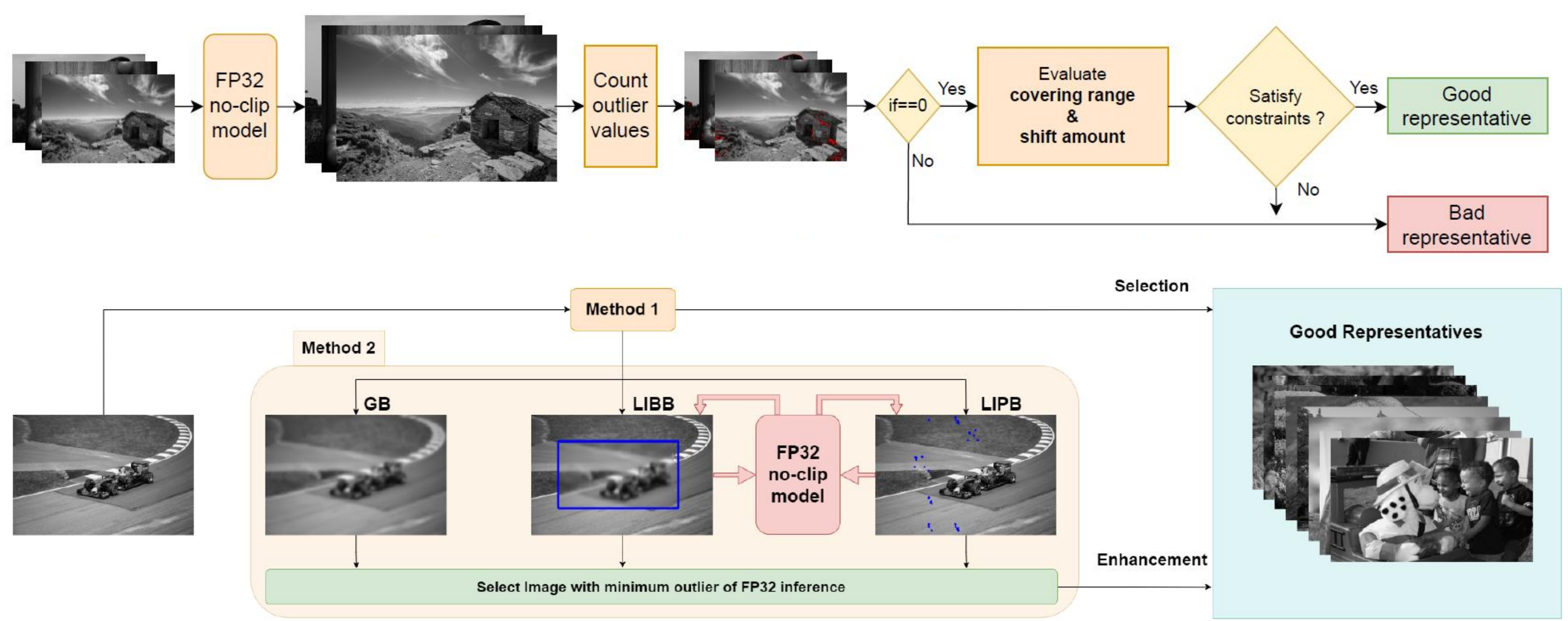
Therefore, if we reduce the outliers, we may also decrease ΔR=Gray and increase the PTQ stability. **This also means that we can understand whether a representative image would be considered as good or bad without performing PTQ. Then, if we correctly identify and use good representative images, we can omit the clipped activation layer without any loss in the INT8 model.**

Procedure

We propose a **pipeline consisting of two methods** to augment RD images in order to be able to apply PTQ to mobile super-resolution networks without using clipped activations:

Method 1 (M1): *Selection of good representative images*

Method 2 (M2): *Enhancing the bad images*



GB: Global Blur	LIBB: Local Iterative Box Blur	LIPB: Local Iterative Point Blur

if $ON(Image) \neq 0$ **then**
$Image \Leftarrow GB(Image)$
end if
	while $ON(Image) \neq 0$ **or** $IN \neq Thr$ **do**
	$Image \Leftarrow LIBB(Image)$
	$IN \Leftarrow IN + 1$
	end while
		while $ON(Image) \neq 0$ **or** $IN \neq Thr$ **do**
		$Image \Leftarrow LIPB(Image)$
		$IN \Leftarrow IN + 1$
		end while