Exploring Semantic Segmentation on the DCT Representation

Shao-Yuan Lo, Hsueh-Ming Hang

National Chiao Tung University

MMAsia 2019 Oral

December 18, 2019



Motivation

- Compressed domain analytics: Perform computer vision tasks (e.g., object classification, detection, tracking, segmentation) in the compressed domain directly.

- No need to perform decoding (reduce computation)
- Reuse encoded information
- Potentially decrease the complexity of computer vision systems (our research goal)

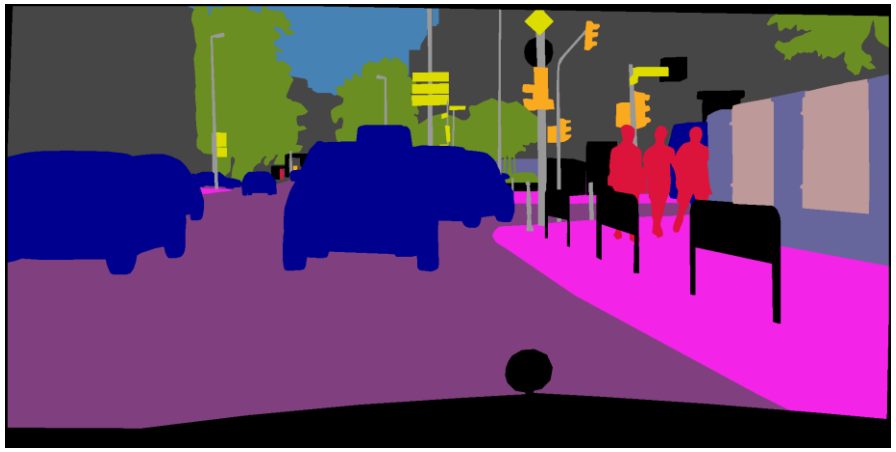
Motivation

- Semantic segmentation: Pixel-level predictions.



Input: RGB image

DL model



Output: Segmentation result

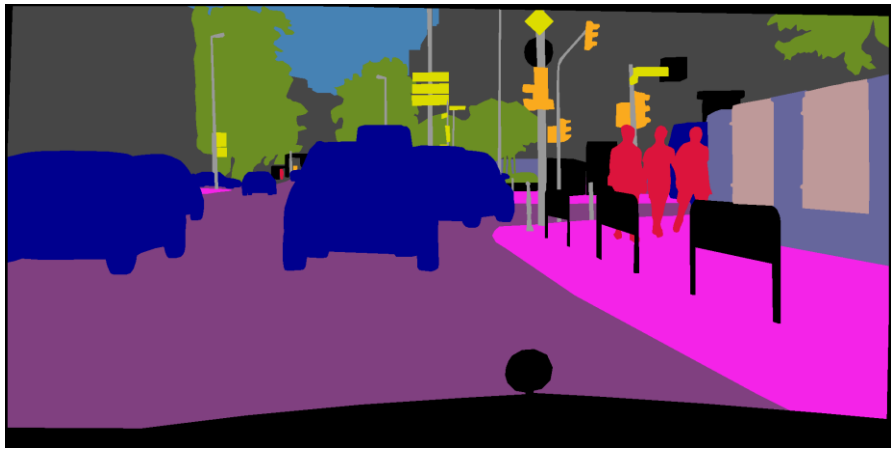
Motivation

- Extract features from compressed representations.



Input: Compressed data
e.g., JPEG (DCT coef.)

DL model for
compressed data



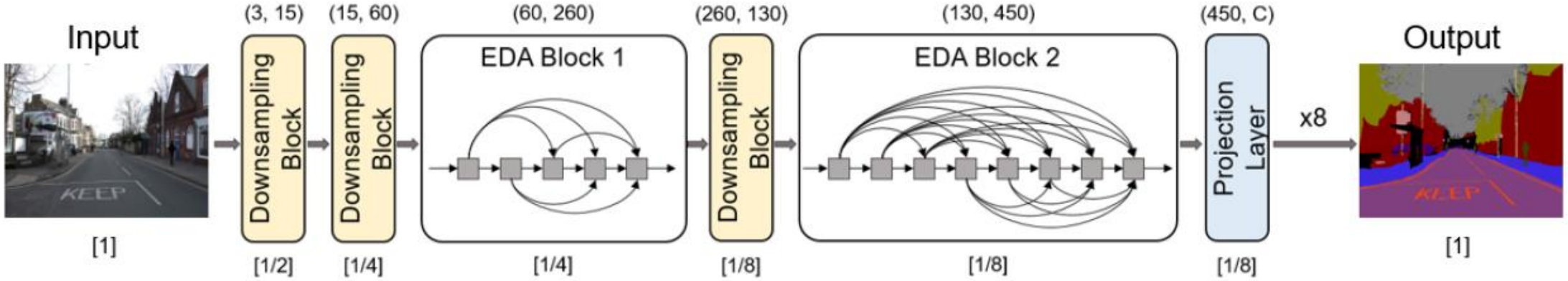
Output: Segmentation results

JPEG Compression

- Convert the color space from RGB to YCbCr.
- Perform block-wise (8×8 pixels) DCT.
- Quantize the DCT coefficients by a quantization matrix.
- Encode the coefficients by entropy encoding.

Network

- EDANet [Lo et al. 2019]
- A CNN for semantic segmentation
- High efficiency and low complexity



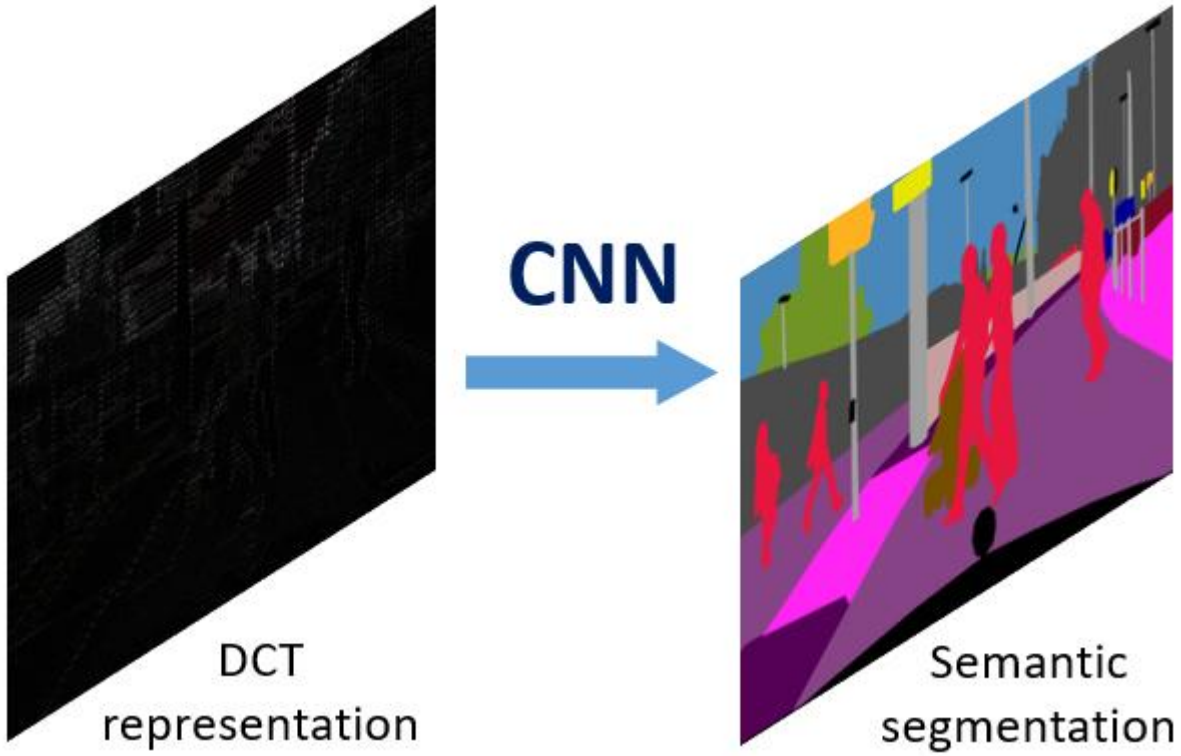
Dataset

- Cityscapes [Cordts et al. 2016]
- 19 classes (road, car, person, building, traffic sign, etc.)
- 5000 images
- Resolution: 1024 x 2048



DCT coefficients

- Take the DCT coefficients of images as the inputs of a CNN.



DCT coefficients

- CNN can use DCT coefficients to do segmentation but get lower accuracy.

Input	mIoU (%)
RGB	63.7*
DCT	59.3

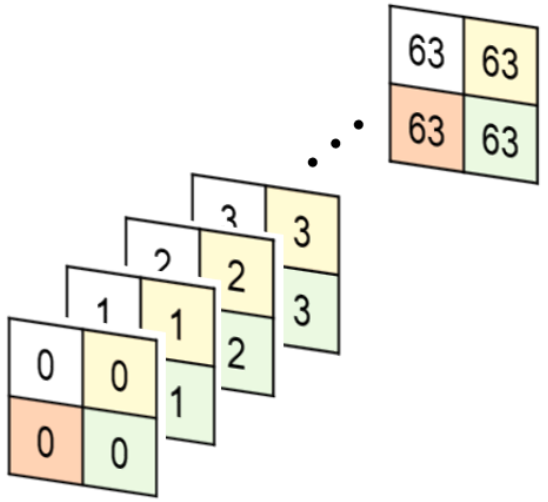
*Models are trained with one-stage and through just 2/3 number of iterations compared to that in EDANet [16] since we compare the relative accuracy in our analysis.

Frequency Component Rearrangement

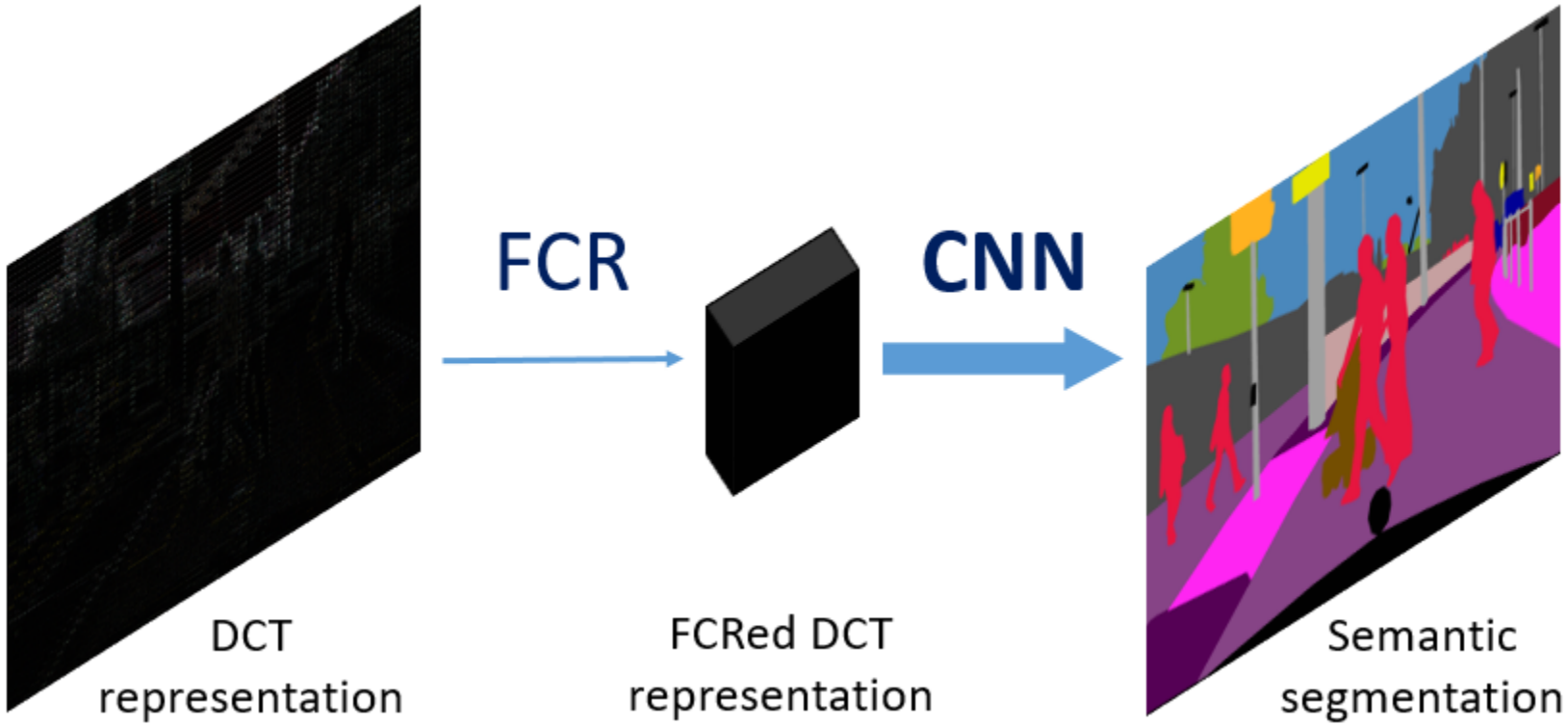
- Rearrange frequency information on the 3rd dimension.
- Input tensor: 512 x 1024 x 3 -> 64 x 128 x (64x3)

0	1	5	6	14	15	27	28	0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42	2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43	3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53	9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54	10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60	20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61	21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63	35	36	48	49	57	58	62	63
0	1	5	6	14	15	27	28	0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42	2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43	3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53	9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54	10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60	20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61	21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63	35	36	48	49	57	58	62	63

FCR



Frequency Component Rearrangement



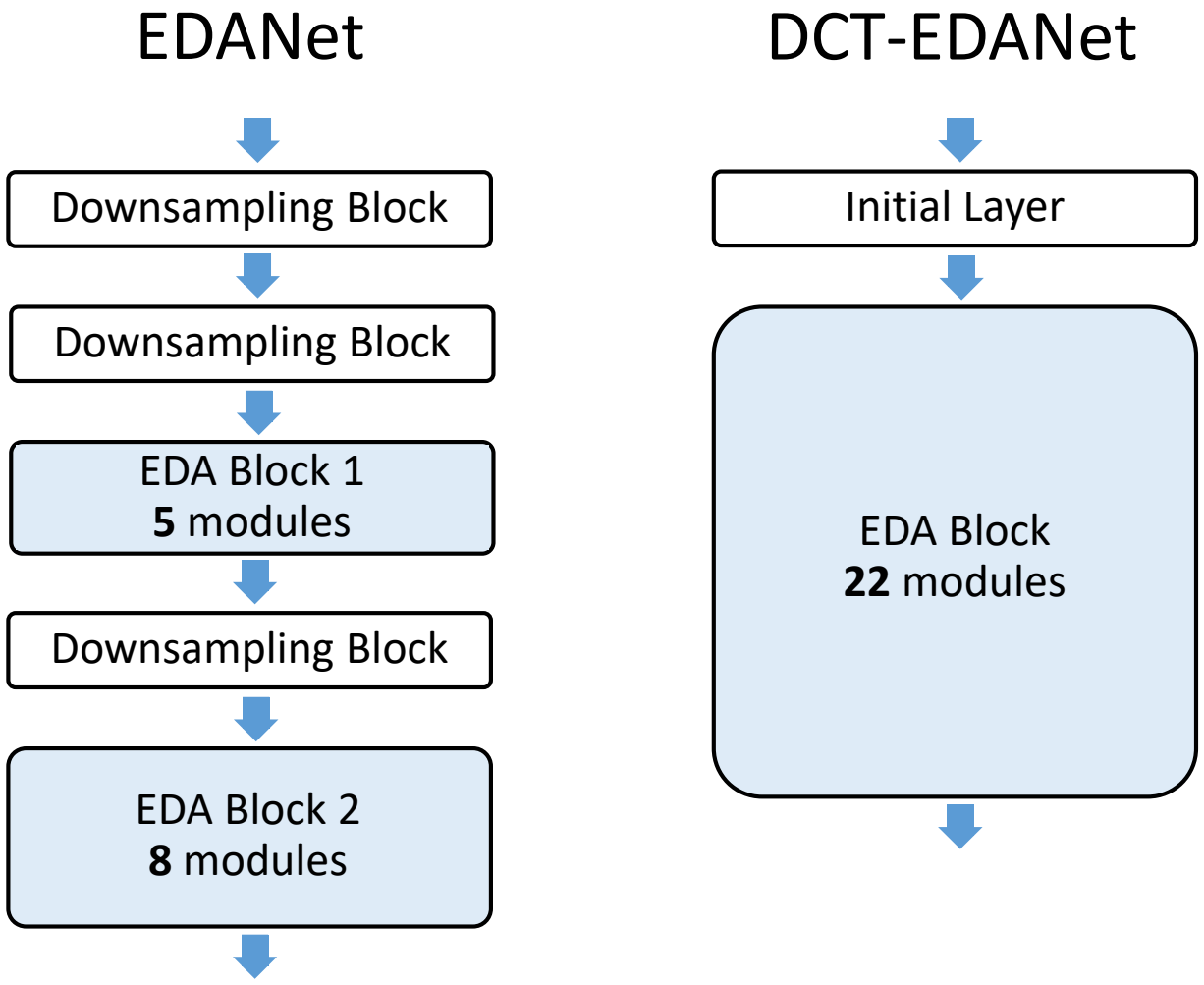
Frequency Component Rearrangement

- The accuracy drops a lot.

Input	mIoU (%)
RGB	63.7
DCT	59.3
FCRed DCT	37.8

DCT-EDANet

- Remove downsamplings
- Increase depths



DCT-EDANet

- DCT-EDANet obtains a dramatic improvement.

Architecture	Input	mIoU (%)	Multi-Adds
EDANet	RGB	63.7	8.97B
EDANet	DCT	59.3	8.97B
EDANet	FCRed DCT	37.8	0.20B
DCT-EDANet	**FCRed DCT**	**61.6**	**8.52B**

DCT-EDANet

- Take the first 16 low-frequency coefficients of each 8×8 block as inputs.
- The accuracy gap between EDANet-DCT and DCT-EDANet is widened from 2.3% to 4.0%, which indicates DCTEDANet are more favorable when the inputs are condensed

$$\begin{bmatrix} -415 & -30 & -61 & 27 & 56 & -20 & -2 & 0 \\ 4 & -22 & -61 & 10 & 13 & -7 & -9 & 5 \\ -47 & 7 & 77 & -25 & -29 & 10 & 5 & -6 \\ -49 & 12 & 34 & -15 & -10 & 6 & 2 & 2 \\ 12 & -7 & -13 & -4 & -2 & 2 & -3 & 3 \\ -8 & 3 & 2 & -6 & -2 & 1 & 4 & 2 \\ -1 & 0 & 0 & -2 & -1 & -3 & 4 & -1 \\ 0 & 0 & -1 & -4 & -1 & 0 & 1 & 2 \end{bmatrix}$$

Input	mIoU (%)
EDANet-DCT-1/4coef	55.0
DCT-EDANet-1/4coef	**59.0**

Frequency Component Selection

- Different combinations of DCT coefficients as inputs.
- Purpose: Discover important coefficients so that we can take only these as inputs.

Y

$$\begin{bmatrix} -415 & -30 & -61 & 27 & 56 & -20 & -2 & 0 \\ 4 & -22 & -61 & 10 & 13 & -7 & -9 & 5 \\ -47 & 7 & 77 & -25 & -29 & 10 & 5 & -6 \\ -49 & 12 & 34 & -15 & -10 & 6 & 2 & 2 \\ 12 & -7 & -13 & -4 & -2 & 2 & -3 & 3 \\ -8 & 3 & 2 & -6 & -2 & 1 & 4 & 2 \\ -1 & 0 & 0 & -2 & -1 & -3 & 4 & -1 \\ 0 & 0 & -1 & -4 & -1 & 0 & 1 & 2 \end{bmatrix}$$

Cb

$$\begin{bmatrix} -415 & -30 & -61 & 27 & 56 & -20 & -2 & 0 \\ 4 & -22 & -61 & 10 & 13 & -7 & -9 & 5 \\ -47 & 7 & 77 & -25 & -29 & 10 & 5 & -6 \\ -49 & 12 & 34 & -15 & -10 & 6 & 2 & 2 \\ 12 & -7 & -13 & -4 & -2 & 2 & -3 & 3 \\ -8 & 3 & 2 & -6 & -2 & 1 & 4 & 2 \\ -1 & 0 & 0 & -2 & -1 & -3 & 4 & -1 \\ 0 & 0 & -1 & -4 & -1 & 0 & 1 & 2 \end{bmatrix}$$

Cr

$$\begin{bmatrix} -415 & -30 & -61 & 27 & 56 & -20 & -2 & 0 \\ 4 & -22 & -61 & 10 & 13 & -7 & -9 & 5 \\ -47 & 7 & 77 & -25 & -29 & 10 & 5 & -6 \\ -49 & 12 & 34 & -15 & -10 & 6 & 2 & 2 \\ 12 & -7 & -13 & -4 & -2 & 2 & -3 & 3 \\ -8 & 3 & 2 & -6 & -2 & 1 & 4 & 2 \\ -1 & 0 & 0 & -2 & -1 & -3 & 4 & -1 \\ 0 & 0 & -1 & -4 & -1 & 0 & 1 & 2 \end{bmatrix}$$

Model	# input coef.	# Y coef.	# Cb coef.	# Cr coef.	mIoU (%)
DCT-EDANet	192	64	64	64	61.6
M-64-0-0	64	64	0	0	59.8
M-49-9-9	67	49	9	9	60.6
M-36-16-16	**68**	**36**	**16**	**16**	**61.2**
M-25-25-25	75	25	25	25	59.7
M-16-16-16	48	16	16	16	59.0
M-16-4-4	**24**	**16**	**4**	**4**	**59.9**
M-16-1-1	18	16	1	1	57.4
M-9-4-4	**17**	**9**	**4**	**4**	**58.7**
M-0-0-16	16	0	0	16	46.4

Frequency Component Selection

- We found the best input component proportion is around **50:25:25**, providing a guideline for future studies on DCT-domain analytics.

- This result is consistent with a principle of the JPEG compression algorithm, in which the chroma information is less critical and thus subsampled in the JPEG codec.

Quantization

- In the JPEG codec, if the compression is lossy, the quantization step is included.

- Compare DCT coefficients quantized by different Q-factors and their corresponding decompressed RGB images.

Quantization

- The proposed method can tolerate serious quantization errors.

Model	Quality factor	mIoU (%)
DCT-EDANet	No	61.6
M-QF70	70	60.5
M-QF50	50	60.6
M-QF30	30	60.0

Conclusion

- To our knowledge, this paper is the first to explore semantic segmentation on the DCT representation.
- We rearrange the DCT coefficients by using FCR. Then, we modify EDANet by discarding all the downsampling operations and deepening the network to maintain the network capacity.
- The elaborated analysis of DCT coefficient selections provides a guideline for future studies on compressed-domain analytics

Thanks for your attention