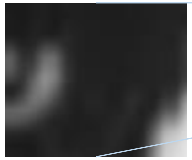
Deep learning for intra prediction: context-adaptive neural networks

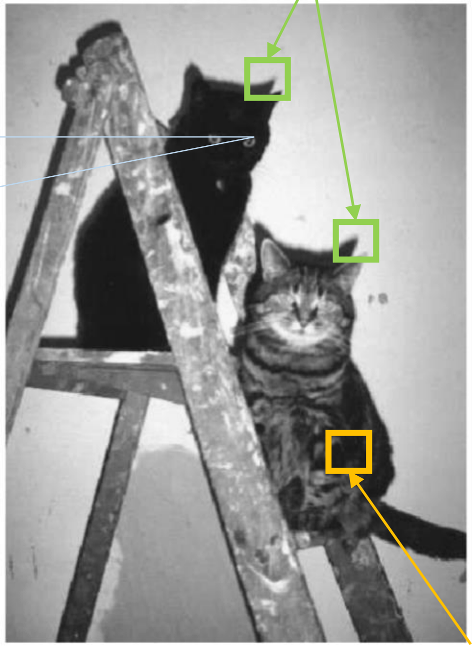
Thierry Dumas, Aline Roumy, Christine Guillemot

INRIA Rennes, France

Fundamentals of image compression

long range spatial dependencies

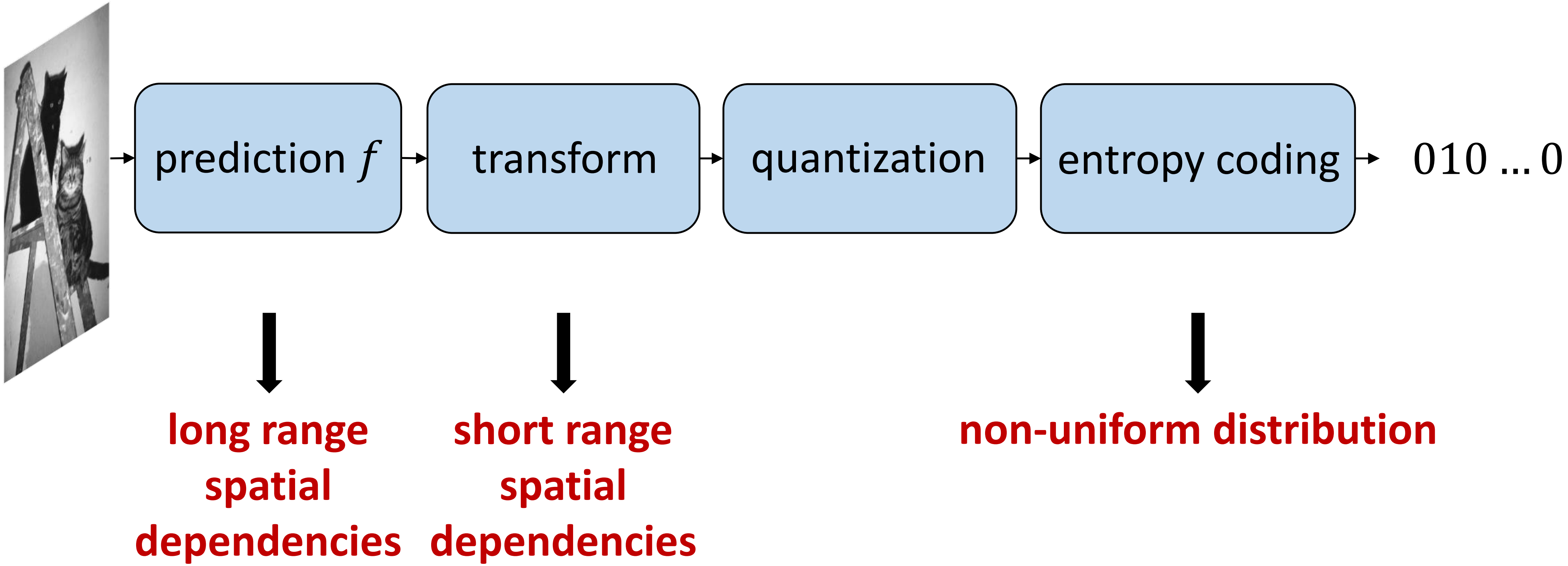




gray levels with various frequencies
→ non-uniform distribution

short range spatial dependencies

Fundamentals of image compression



From optimal prediction to feasible prediction

- **optimality**

  observed random variables $\{X_{n,0}, \dots, X_{n,l-1}\}$, unobserved random variable S_n

  optimal prediction $\hat{S}_n{}^{*} = \mathbb{E}\left[S_n | X_{n,0}, \dots, X_{n,l-1}\right]$ ➡ **law costly to transmit**

- **practice**

  a. define a finite set of laws $\{f_0, \dots, f_{p-1}\}$
  b. find $f = f_i$ on the encoder side
  c. send i from the encoder to the decoder

But
- $\boldsymbol{p}$ ↗, **transmission cost** ↗
- $\{\boldsymbol{f}_{\mathbf{0}}, \dots, \boldsymbol{f}_{\boldsymbol{p}-\mathbf{1}}\}$ **linear, representing simple dependencies**

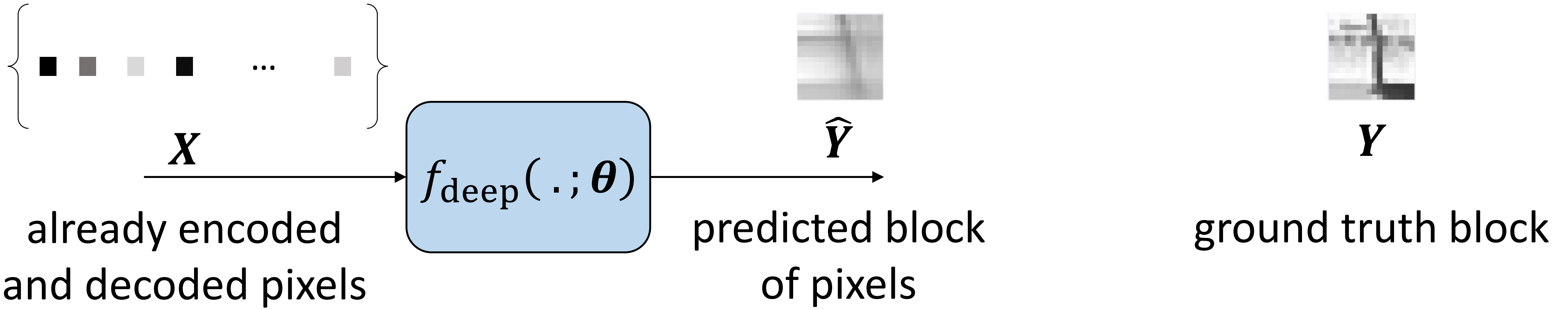
Goal of our work

prediction function f_{deep}

- modeling complex dependencies between pixels
- predicting large unknown regions of pixels

learning deep neural network $f_{\text{deep}}(\,.\,;\boldsymbol{\theta})$

I – Generic neural network based intra prediction

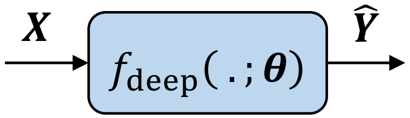


prediction MSE

$$\boldsymbol{\theta}^{*} = \min_{\boldsymbol{\theta}} \mathbb{E}\left[\left\|\boldsymbol{Y} - f_{\text{deep}}(\boldsymbol{X};\boldsymbol{\theta})\right\|_2^2 + \lambda \|[\boldsymbol{\theta}]_W\|_2^2\right], \lambda \in \mathbb{R}_+^{\;*}$$

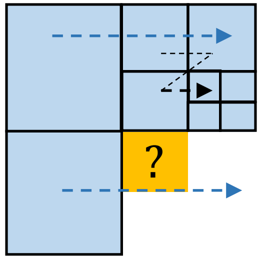
weights regularization

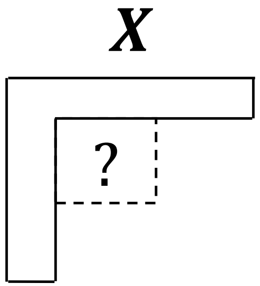
II – Challenges of building a deep predictor



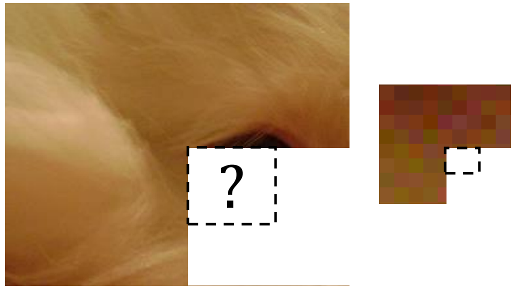
- shape of $\boldsymbol{X}$?

raster scanning/Z-scanning

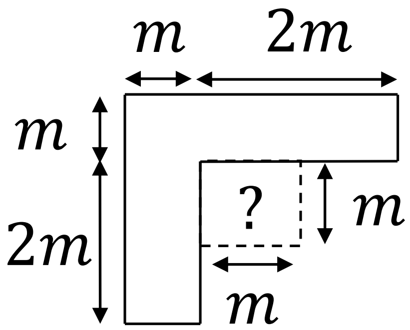




- size of $\boldsymbol{X}$?



ground truth:

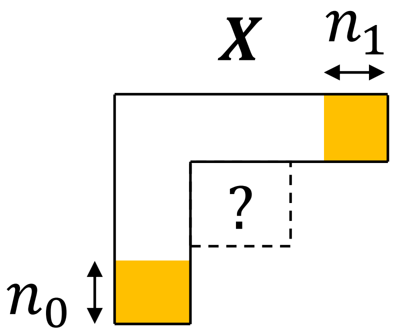


size growing linearly

II – Challenges of building a deep predictor

- variable number of available neighboring pixels?

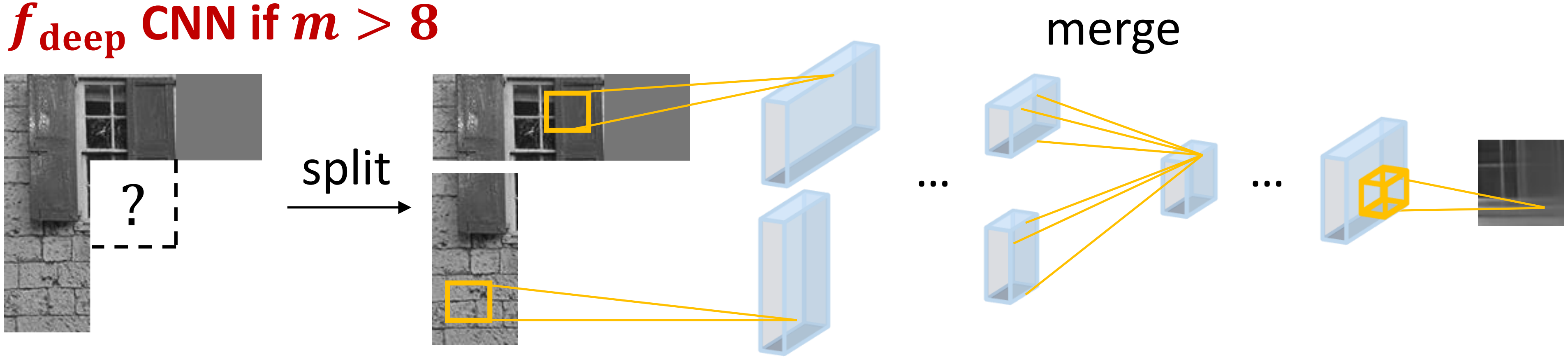
avoid one training per case



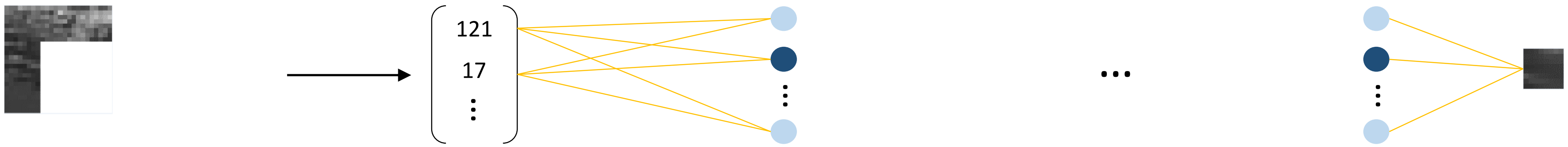
$\boldsymbol{n_0, n_1 \sim \mathcal{U}[\![1, m]\!]}$
during the training

II – Challenges of building a deep predictor

- architecture of f_{deep}?
  - **f_{deep} CNN if $m > 8$**



  - **f_{deep} fully-connected otherwise**

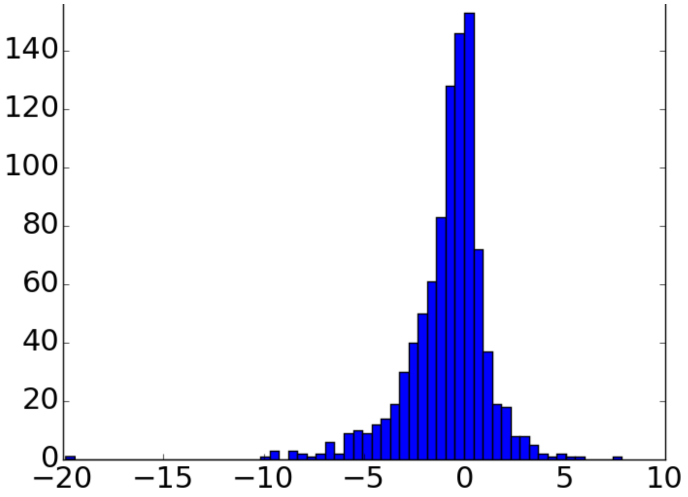


III – Success rate of the neural networks

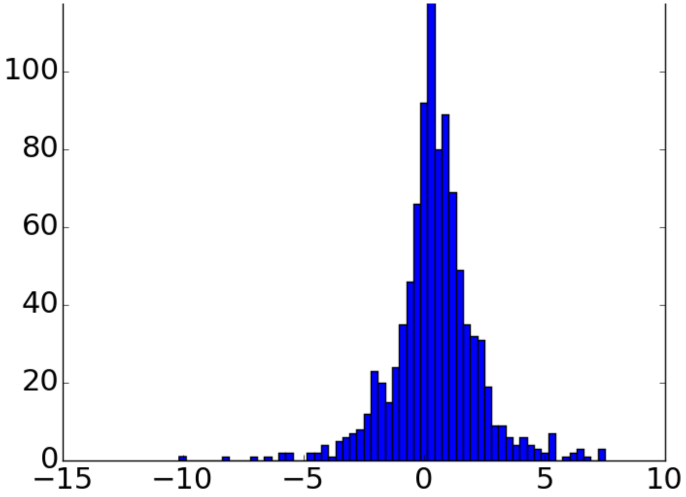
baseline: best prediction function f_{HEVC} among the 35 HEVC functions.

difference in prediction PSNR

8×8 blocks

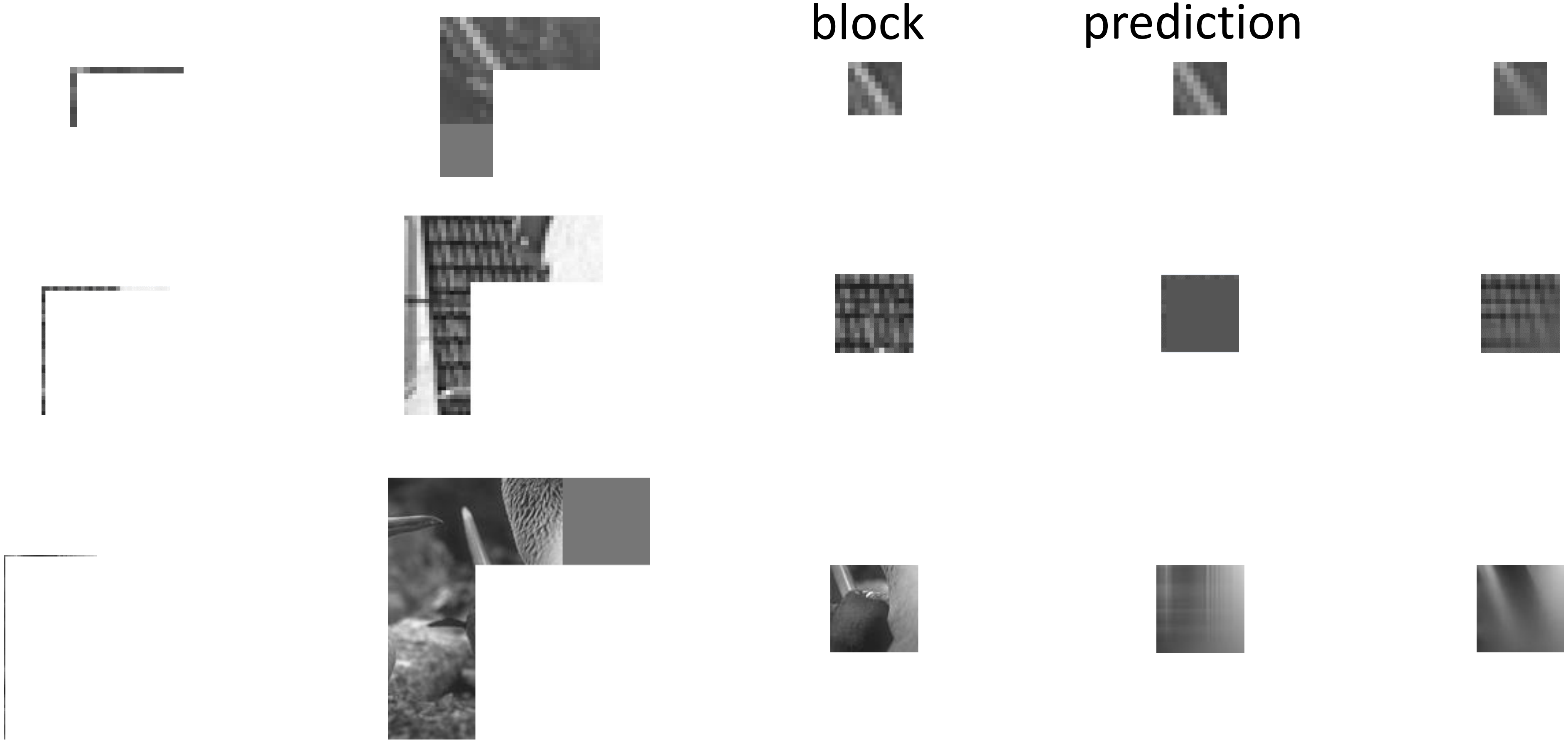


64×64 blocks



% wins f_{deep} over f_{HEVC}: **32%** **68%**

IV – Quality of prediction



V – Performance in terms of rate-distortion

signalling → neural network function **1**

signalling → 35 HEVC intra prediction functions **0** and 2 to 5 bits

bitrate savings of the neural networks w.r.t HEVC for **same distortion**

video first luminance frame	bitrate saving
Traffic 2560×1600	3.76%
BQTerrace 1920×1080	2.44%
BasketballDrive 1920×1080	**5.20%**
Cactus 1920×1080	3.05%
ParkScene 1920×1080	2.58%
Kimono 1920×1080	2.92%
BQSquare 416×240	**2.21%**

IV – Conclusion

+ learning an intra prediction function modelling complex dependencies between pixels

+ adapting the function w.r.t:
  - the size of the block to be predicted
  - the number of available neighboring pixels

- computation time

- storage of the neural network parameters (10 million approximatively).

Thanks you for your attention!

For further details,

https://www.irisa.fr/temics/demos/prediction_neural_network/PredictionNeuralNetwork.htm