LISIC - Webinar

Toward a noise perception model for photorealistic image synthesis

Jérôme Buisine (PhD Student)
Supervisors: Christophe Renaud and Samuel Delepoulle
Team: IMAP (Images et Apprentissage)

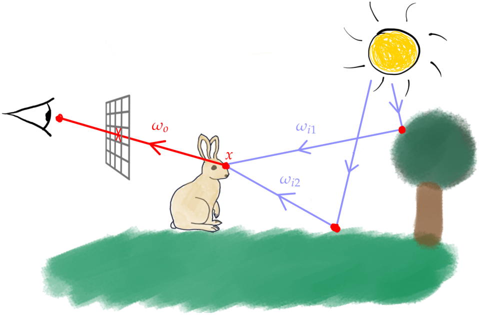
December, 10 2020

* ANR support : project ANR-17-CE38-0009
Univ. Littoral Côte d'Opale, LISIC, F-62100 Calais, France

Agenda

Context

Context



$$L_o(x, \omega_o) = L_e(x, \omega_o) + \int_{\Omega} L_i(x, \omega_i) \cdot f_r(x, \omega_i \rightarrow \omega_o) \cdot \cos\theta_i d\omega_i \qquad (1)$$

Photorealistic image synthesis

- Global illumination rendering
- Monte Carlo

Context: noise in photorealistic image

(a) After 1 sample

(b) After 20 samples

(c) After 10, 000 samples

Context: noise in photorealistic image

(a) After 1 sample

(b) After 20 samples

(c) After 10, 000 samples

Question:

How can human perceive this MC noise ?

Dataset

Dataset creation: need of human data

Problem of photorealistic image synthesis rendering

- No-reference context during rendering
- Unavailable models for noise perception in MC generated images
- No human perceptual reference data

Dataset creation: need of human data

Problem of photorealistic image synthesis rendering

- No-reference context during rendering
- Unavailable models for noise perception in MC generated images
- No human perceptual reference data

A solution

Collect human subjective perceptual threshold during rendering as ground truth

Dataset creation: need of human data

Problem of photorealistic image synthesis rendering

- No-reference context during rendering
- Unavailable models for noise perception in MC generated images
- No human perceptual reference data

A solution

Collect human subjective perceptual threshold during rendering as ground truth

Build a model

Use these perceptual thresholds into a perceptual noise model

Perception: definition

Just-Noticeable Difference (JND)

Noise can be viewed as a perceptible difference into image

20 samples

1000 samples

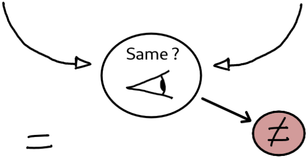
Perception: definition

Just-Noticeable Difference (JND)

Noise can be viewed as a perceptible difference into image

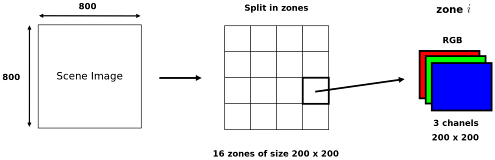
20 samples

1000 samples



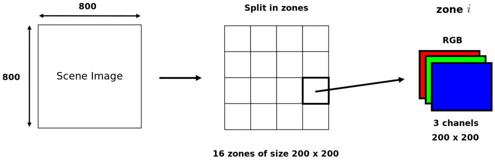
Dataset creation: collect human subjective threshold

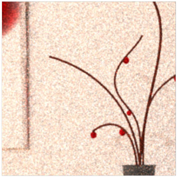
Our way of getting perceptual subjective thresholds



Dataset creation: collect human subjective threshold

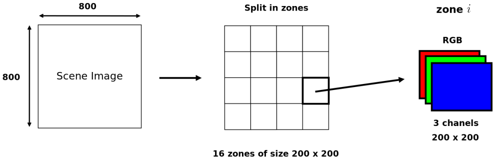
Our way of getting perceptual subjective thresholds

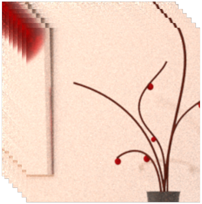




Dataset creation: collect human subjective threshold

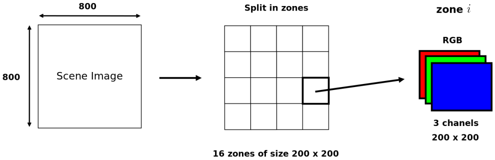
Our way of getting perceptual subjective thresholds

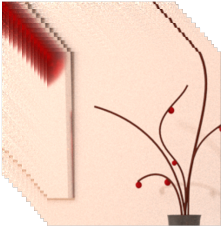




Dataset creation: collect human subjective threshold

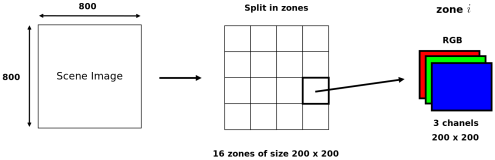
Our way of getting perceptual subjective thresholds

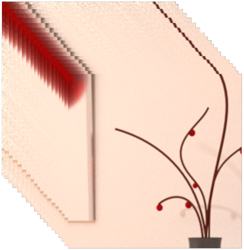




Dataset creation: collect human subjective threshold

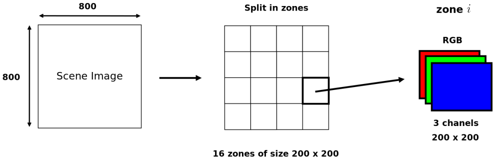
Our way of getting perceptual subjective thresholds

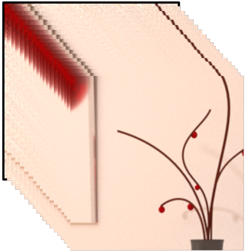


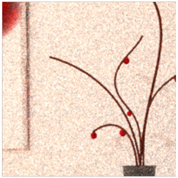


Dataset creation: collect human subjective threshold

Our way of getting perceptual subjective thresholds





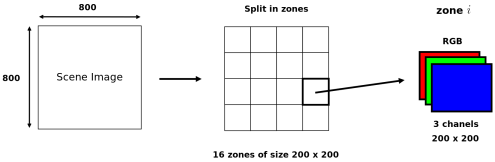


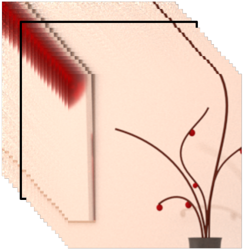
20 samples

3000 samples (reference)

Dataset creation: collect human subjective threshold

Our way of getting perceptual subjective thresholds



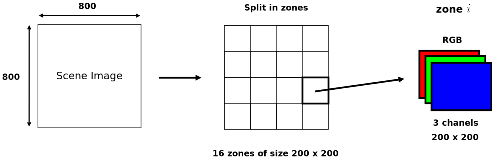


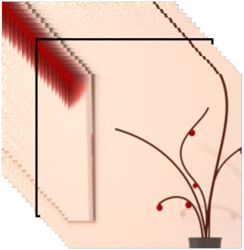
220 samples

3000 samples (reference)

Dataset creation: collect human subjective threshold

Our way of getting perceptual subjective thresholds



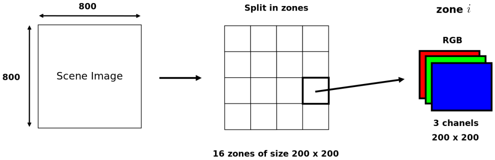


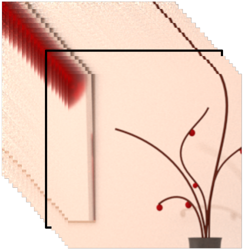
500 samples

3000 samples (reference)

Dataset creation: collect human subjective threshold

Our way of getting perceptual subjective thresholds



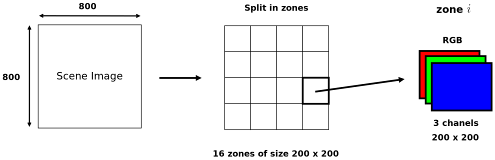


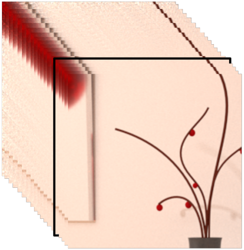
900 samples

3000 samples (reference)

Dataset creation: collect human subjective threshold

Our way of getting perceptual subjective thresholds



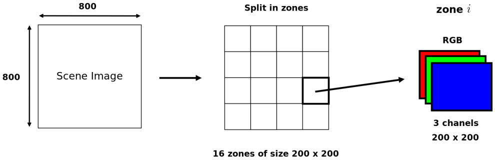


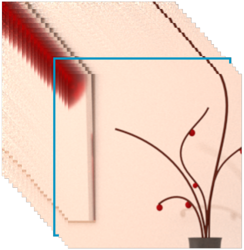
1400 samples

3000 samples (reference)

Dataset creation: collect human subjective threshold

Our way of getting perceptual subjective thresholds



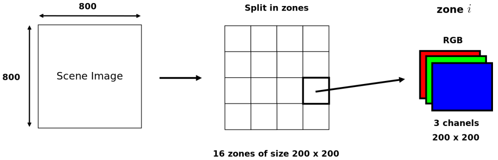


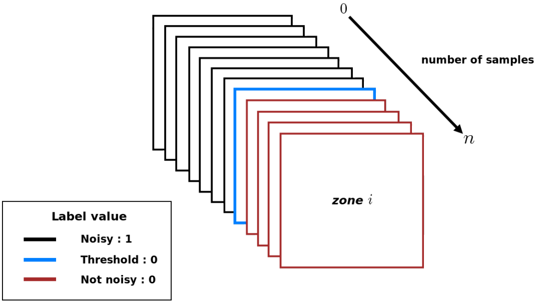
1400 samples

3000 samples (reference)

Dataset creation: collect human subjective threshold

Our way of getting perceptual subjective thresholds





Dataset creation: overview

313	312	274	271
310	301	308	235
248	292	222	240
211	151	139	177

(a) Human thresholds (Mean Opinion Score)

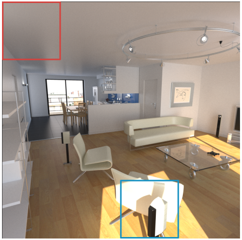
(b) Human reference

(c) After 900 samples

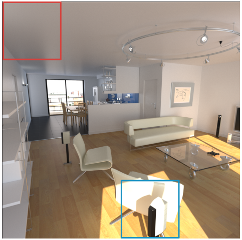
Dataset creation: overview

313	312	274	271
310	301	308	235
248	292	222	240
211	151	139	177

(a) Human thresholds (Mean Opinion Score)



(b) Human reference

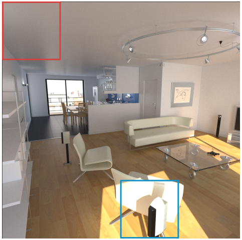


(c) After 900 samples

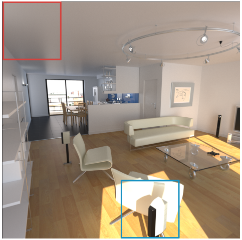
Dataset creation: overview

313	312	274	271
310	301	308	235
248	292	222	240
211	151	139	177

(a) Human thresholds (Mean Opinion Score)



(b) Human reference
SSIM: 0.70 (< 0.95)



(c) After 900 samples
SSIM: 1

Structural Similarity Index (SSIM)

SSIM metric quantifies the visibility of errors between a distorted image and a reference image using a variety of known properties of the human visual system.

Build of new dataset

Previous dataset

- 9 viewpoints from scenes
- different renderers (maxwell, igloo, cycle...)
- hence, different algorithms

Build of new dataset

Previous dataset

- 9 viewpoints from scenes
- different renderers (maxwell, igloo, cycle...)
- hence, different algorithms

New dataset

- 40 viewpoints with $10,000$ images of 1 sample (HD images)
- only **pbrt-v3** renderer
- use of **path-tracing**
- available soon

Build of new dataset

Why saving image of 1 sample ?

- generate $\binom{10000}{k}$ images of k samples from pool of $10,000$ samples

Build of new dataset

Why saving image of 1 sample ?

- generate $\binom{10000}{k}$ images of k samples from pool of $10,000$ samples
  $\Rightarrow \binom{10000}{20} \approx 4.3\mathrm{e}61$

Build of new dataset

Why saving image of 1 sample ?

- generate $\binom{10000}{k}$ images of k samples from pool of $10,000$ samples
  $\Rightarrow \binom{10000}{20} \approx 4.3\mathrm{e}61$
- **posterior** study of samples **distribution**

Build of new dataset

Why saving image of 1 sample ?

- generate $\binom{10000}{k}$ images of k samples from pool of $10,000$ samples
  $\Rightarrow \binom{10000}{20} \approx 4.3\mathrm{e}61$
- **posterior** study of samples **distribution**
- use of deep learning approach (RNN, GAN, Autoencoder...)

Build of new dataset

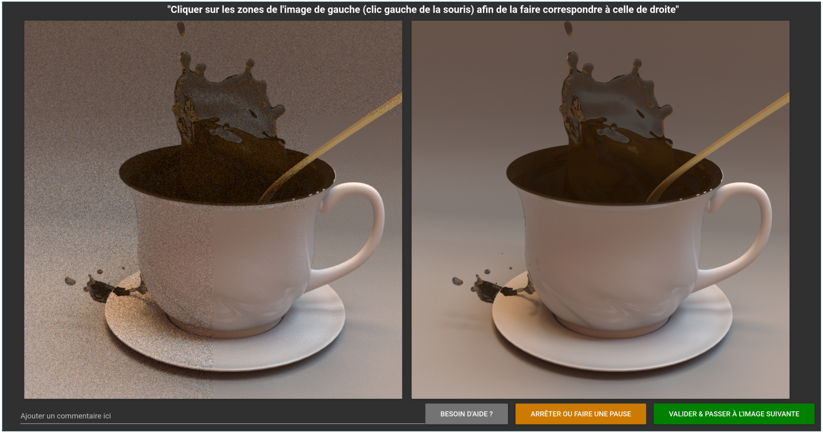


Figure 5: SIN3D web application

Expected model

Binary classification

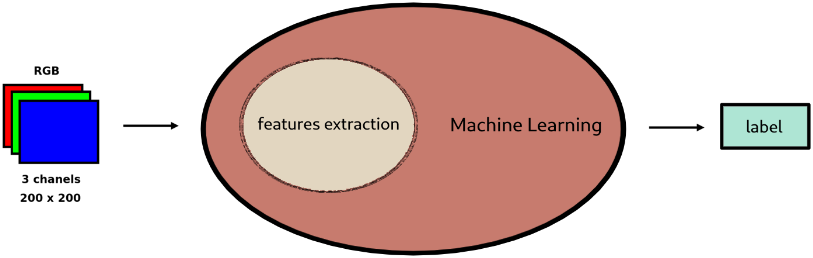
- Model which labels image as **noisy** or **not** (**converged** or **not**)
- Supervised learning

Expected model

Binary classification

- Model which labels image as **noisy** or **not** (**converged** or **not**)
- Supervised learning

Common pipeline used



Why this kind of model ?

Why this kind of model ?

- **stopping criterion** during rendering based on sub-blocks of rendered image

Why this kind of model ?

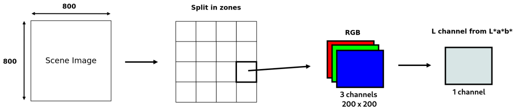
- **stopping criterion** during rendering based on sub-blocks of rendered image
- **save** computation time

Why this kind of model ?

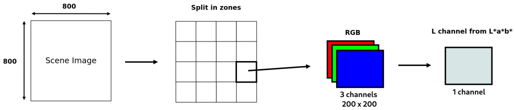
- **stopping criterion** during rendering based on sub-blocks of rendered image
- **save** computation time
- target more complex parts of the scene

Noise detection

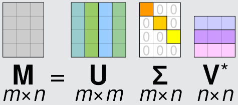
SVD attributes



SVD attributes



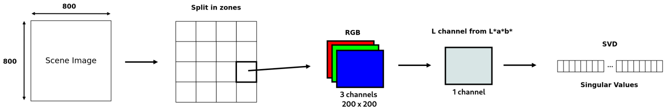
Singular Value Decomposition



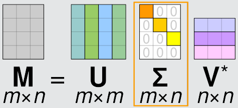
where:

- M is an $m \times n$ real or complex matrix
- U is an $m \times m$ real or complex unitary matrix.
- Σ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal.
- V is an $n \times n$ real or complex unitary matrix.

SVD attributes



Singular Value decomposition

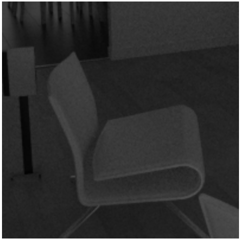


where:

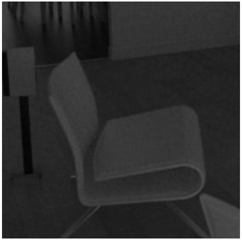
- M is an $m \times n$ real or complex matrix
- U is an $m \times m$ real or complex unitary matrix.
- Σ is an $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal.
- V is an $n \times n$ real or complex unitary matrix.

SVD attributes

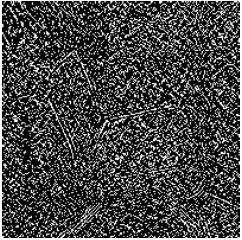
Possibility to decompose image using SVD into structure dependent and non-dependent images (Wang et al. 2013).



(a) L chanel (500 samples)

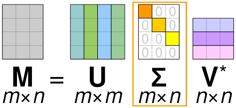


(b) SVD reconstruction (0, 50)



(c) SVD reconstruction (50, 200)

SVD-Entropy and RNN



Shannon entropy of singular values can be defined as SVD-Entropy (O.Alter, P.O.Brown, and D.Bolstein 2000):

$$H_{SVD} = -\frac{1}{\log(O)} \sum_{i=1}^{O} \overline{\sigma}_i \log(\overline{\sigma}_i) \qquad (2)$$

where :

$$\overline{\sigma}_i = \sigma_i^2 / \sum_{p=1}^{O} \sigma_p^2 \qquad (3)$$

SVD-Entropy

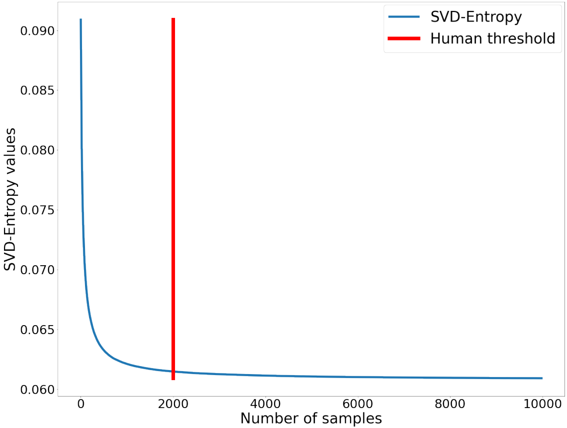
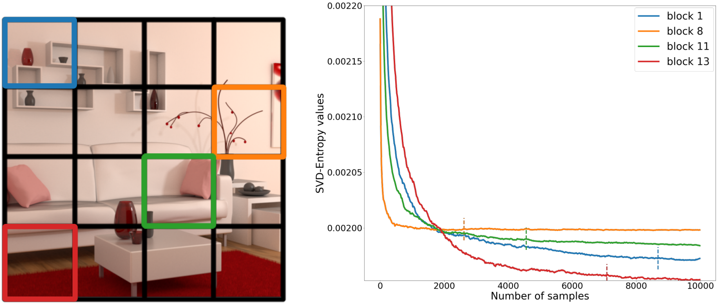


Figure 7: H_{SVD} evolution during over Kitchen image.

SVD-Entropy



(a) Selected blocks indications

(b) Normalized H_{SVD} evolution overview for the 4 selected blocks

Figure 8: Overview of the evolution of H_{SVD} for the *Living room 3* image.

SVD-Entropy and RNN

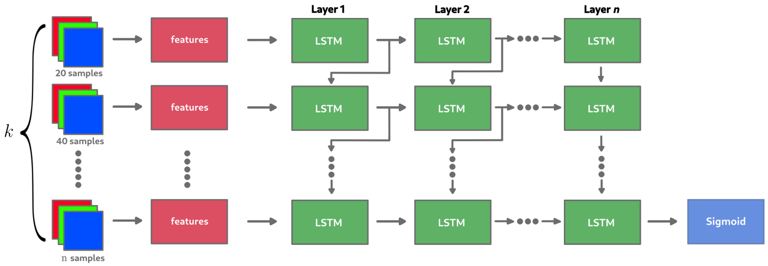


Figure 9: Recurrent neural network with different samples images level as input

SVD-Entropy and RNN

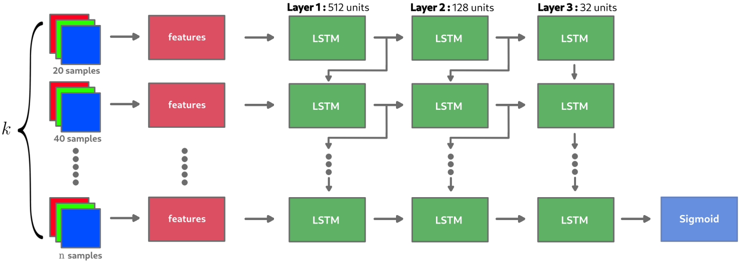


Figure 10: Recurrent neural network with different samples images level as input

SVD-Entropy and RNN

Parameters studied:

- The size of the sequence of RNN with $k \in [3, 4, ..., 10]$;
- $m \in [4, 25, 100, 400]$ (number of sub-blocks cut out within the block). Sub-blocks are respectively of size 100×100, 40×40, 20×20 and 10×10 ;
- Batch size: $b_s \in [64, 128]$;
- Samples sequence step: $n \in [20, 40, 80]$;
- Input normalization: *bnorm* or *snorm* ;
- The value extracted from a sub-block $F \in [H_{SVD}, H^1_{SVD}, H^2_{SVD}]$.

where:

$$H^1_{SVD} = -\frac{1}{\log(\frac{O}{4})} \sum_{i=0}^{O/4} \overline{\sigma}_i \log_2 \overline{\sigma}_i \qquad H^2_{SVD} = -\frac{1}{\log(O - \frac{O}{4})} \sum_{i=O/4}^{O} \overline{\sigma}_i \log_2 \overline{\sigma}_i$$

SVD-Entropy and RNN

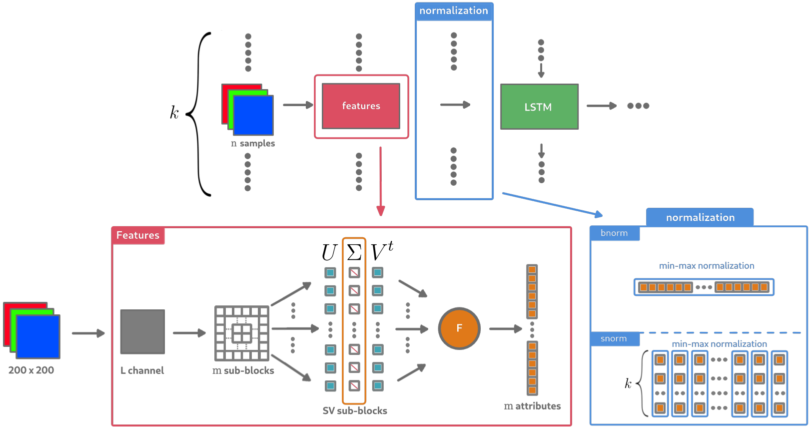


Figure 11: Pipeline for SVD-Entropy and RNN

SVD-Entropy and RNN

Fixed parameters:

- **RNN:** LSTM (512) / LSTM (128) / LSTM (32) / Sigmoid (1) ;
- Dropout for each LSTM layers set to 40% ;
- Recurrent activation function: *Hard Sigmoid* (input, forget, and output gates);
- Activation function: *Sigmoid* (hidden state and output hidden state);
- Balanced samples weights when propagating binary crossentropy loss.

Dataset specifications

Around 300.000 samples (depending of k) obtained from the 40 viewpoints. 12 blocks used as train data set, the 4 others as testing data set part. Same dataset (train / test) is used for each run (parameters combination).

SVD-Entropy and RNN

Comparisons metrics:

- Accuracy: fraction of predictions model got right;
- AUC ROC: Area Under Curve of receiver operating characteristic curve.

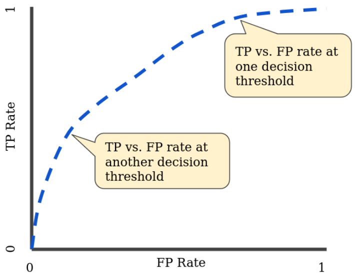


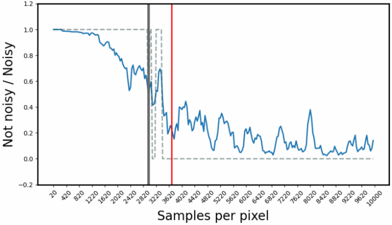
Figure 12: AUC ROC for regression logistic model

SVD-Entropy and RNN

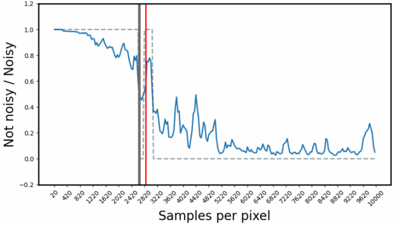
k	m	F	b_s	$bnorm$	$snorm$	n **step**	**Acc Train**	**Acc Test**	**AUC Train**	**AUC Test**
8	100	H_{SVD}	128	0	1	40	84.58 %	82.74 %	84.44 %	82.55 %
5	100	H_{SVD}	128	0	1	80	83.78 %	82.87 %	83.32 %	82.47 %
6	100	H_{SVD}	64	0	1	40	84.18 %	82.61 %	84.01 %	82.45 %
7	100	H_{SVD}	64	0	1	40	84.62 %	82.79 %	84.24 %	82.44 %
7	100	H_{SVD}	128	0	1	40	84.65 %	82.75 %	84.36 %	82.42 %
7	100	H_{SVD}	64	0	1	80	83.67 %	82.36 %	83.70 %	82.38 %
5	100	H_{SVD}	64	0	1	80	83.61 %	82.17 %	83.85 %	82.27 %
9	100	H_{SVD}	64	0	1	40	83.46 %	81.99 %	83.84 %	82.21 %
10	100	H_{SVD}	128	0	1	40	84.52 %	82.58 %	84.20 %	82.16 %

Table 1: 10 best parameters combinations results for RNN model

SVD-Entropy and RNN



(a) Block 5 on Arc sphere viewpoint

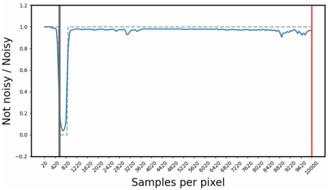


(b) Block 8 on Arc sphere viewpoint

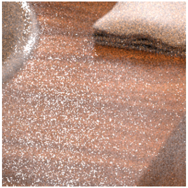
Prediction fluctuation

To overcome this problem and to make thresholds prediction more robust, it was proposed to consider that a block is no longer noisy after **3 successive noiseless** predictions.

SVD-Entropy and RNN



(a) Predictions over the block 10 of *Bathroom* viewpoint



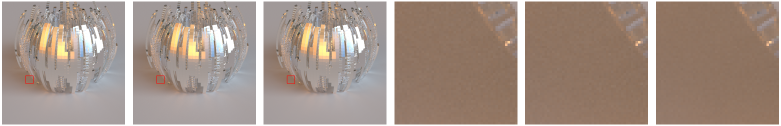
(b) Still noisy block 10 with 500 samples

(c) Reference block 10 with $10,000$ samples

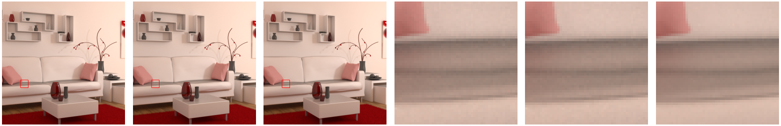
Critical prediction

The block targeted here from the bathroom point of view is still noisy up to $10,000$ samples and seems to contain a significant light reflection.

SVD-Entropy and RNN



SSIM : 0.9840 SSIM : 0.9775 SSIM : 1.0 SSIM : 0.9776 SSIM : 0.9780 SSIM : 1.0



SSIM : 0.9830 SSIM : 0.9903 SSIM : 1.0 SSIM : 0.9927 SSIM : 0.9921 SSIM : 1.0

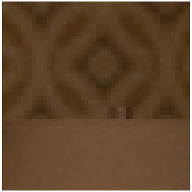
SSIM : 0.9880 SSIM : 0.9604 SSIM : 1.0 SSIM : 0.9878 SSIM : 0.9650 SSIM : 1.0

RNN - H_{SVD} Human Reference RNN - H_{SVD} Human Reference

SVD-Entropy and RNN

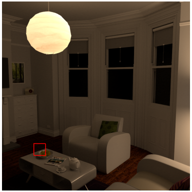
SSIM : 0.9898

SSIM : 1.0

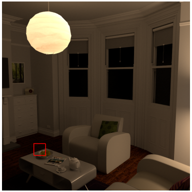


SSIM : 0.9833

SSIM : 1.0

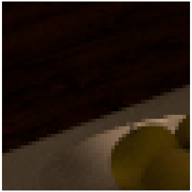


SSIM : 0.9845



SSIM : 1.0

SSIM : 0.9806



SSIM : 1.0

RNN - H_{SVD}	Reference	RNN - H_{SVD}	Reference

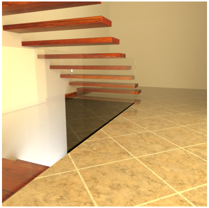
SVD-Entropy and RNN

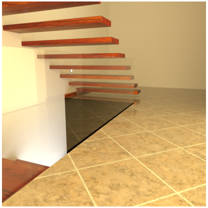
San miguel

10000	8220	6540	10000
9460	8500	8420	10000
6620	8540	10000	10000
9380	6620	10000	9620

Staircase 2

3180	2820	4860	7300
3180	3820	5900	9460
5620	3660	2260	2820
2500	2220	1780	2020





Point of view | Predicted thresholds | H_{SVD} RNN prediction | 10, 000 samples

Conclusion

Conclusion

- Generic method to establish a perceptual stopping criterion
- Some critical cases (lack of data for better generalisation ?)
- Data are available in `https://prise3d.univ-littoral.fr`

Conclusion

Future works:

- MDPI Entropy journal : SVD-Entropy and RNN (submitted) ;
- Use of HDR images for same experiment and make comparisons ;
- Application of *Median Of meaNs* in rendering. Conference or graphics-oriented journal (in progress) ;
- Features selection optimisation : Conference or journal oriented in machine learning / optimisation ;
- Image database : human thresholds, images generated with 1 sample (RAWLS) and images in PNG formats.

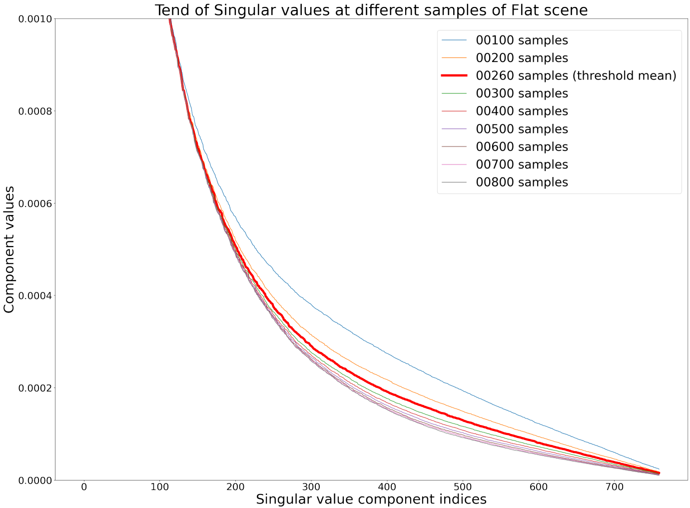
Conclusion

In continuity :

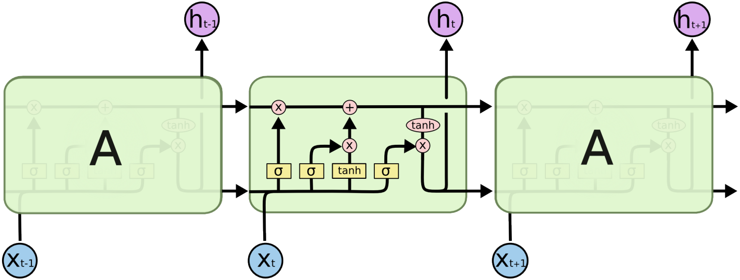
- Improve deep learning works (GAN for denoising) ;
- Create new base with 3D images (stereoscopic).

Thanks for your attention!

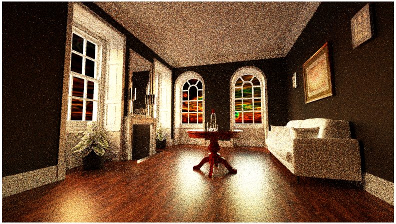
Backup: use of singular values



Backup: LSTM cells



Backup: distribution analysis



(a) Variance

(b) Standard deviation

(c) Skewness

(d) Kurtosis