Deep Learning for Climate

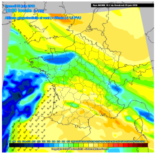
Nicolas Thome - Cnam Paris - CEDRIC / MSDMA
Joint work with V. Le Guen - EDF R&D
January 11, 2019

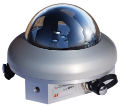
Climate and AI Workshop - Sorbonne Université (SU)

Outline

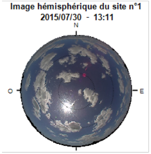
Big data in Climate and Beyond

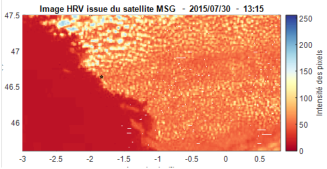
- **Superabundance of data:** times series (sensor measurements), images (fisheye, satellite), spatio-temporal data (weather forecasts), videos, text, *etc*

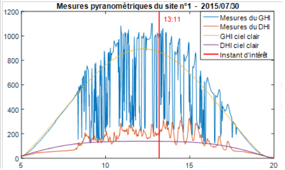




weather forecasts | Sensors, pyranometers | 100M monitoring cameras





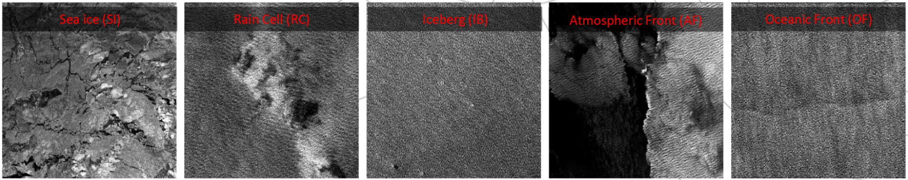


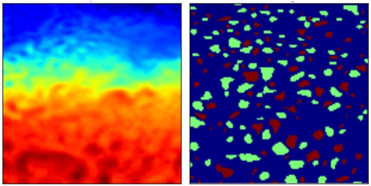
- Obvious need for **Artificial Intelligence** with these data
  $\Rightarrow$ **Recognition, Decision Making**

Decision Making in Climate

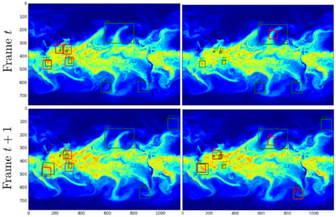
- **Huge number of applications:** classification, *e.g.* RADAR images, segmentation, *e.g.* eddies, forecasting, *e.g.* extreme weather event

[Chen et al., 2018b]





- Anticyclonic
- Cyclonic
- Non eddy

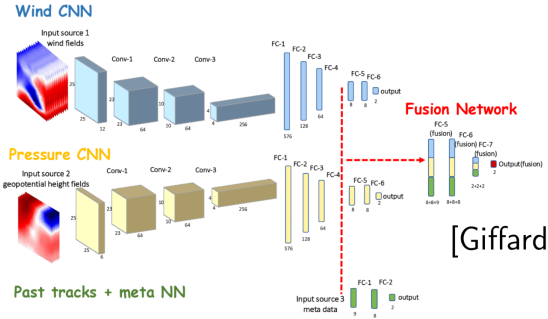


[Lguensat et al., 2018]

[Racah et al., 2017]

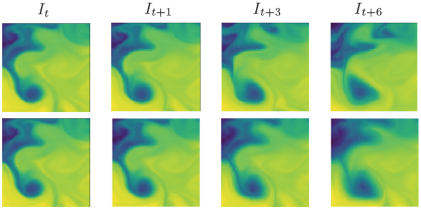
Decision Making in Climate

- Using Mutiple inputs (wind, height, meta-data): hurricane track forecast



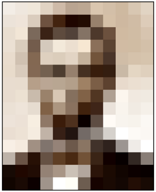
[Giffard-Roisin et al., 2018]

- Exploiting external kowledge: sea surface temperature prediction



[de Bezenac et al., 2018]

Recognition of low-level signals: filling the semantic gap



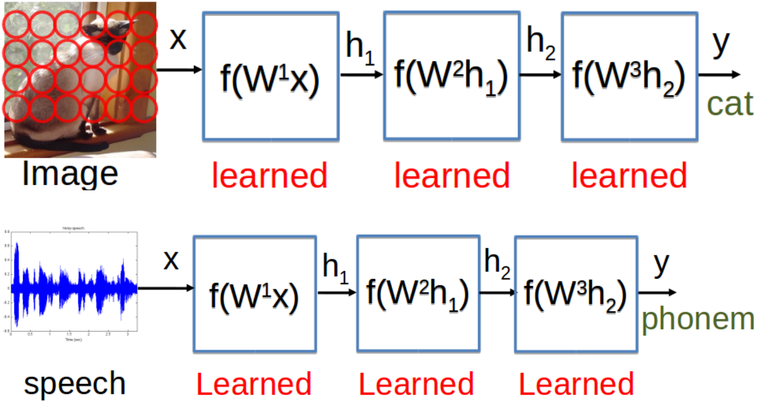
What we perceive *vs*
What a computer sees

243	239	240	225	206	185	188	218	211	206	216	225
242	239	218	110	67	31	34	152	213	206	208	221
243	242	123	58	94	82	132	77	108	208	208	215
235	217	115	212	243	236	247	139	91	209	208	211
233	208	131	222	219	226	196	114	74	208	213	214
232	217	131	116	77	150	69	56	52	201	228	223
232	232	182	186	184	179	159	123	93	232	235	235
232	236	201	154	216	133	129	81	175	252	241	240
235	238	230	128	172	138	65	63	234	249	241	245
237	236	247	143	59	78	10	94	255	248	247	251
234	237	245	193	55	33	115	144	213	255	253	251
248	245	161	128	149	109	138	65	47	156	239	255
190	107	39	102	94	73	114	58	17	7	51	137
23	32	33	148	168	203	179	43	27	17	12	8
17	26	12	160	255	255	109	22	26	19	35	24

- Illumination variations
- View-point variations
- Deformable objects
- intra-class variance
- *etc*

⇒ **Need for "good" Intermediate Representations**

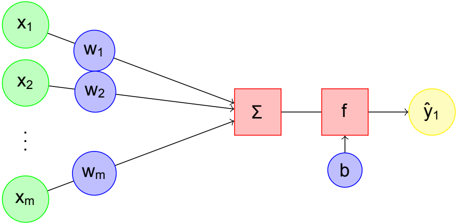
Deep Learning (DL) & Recognition of low-level signals



- **DL: learning intermediate representations**
  - ⊕ **Deep**: hierarchy, gradual learning
  - ⊕ Common learning methodology, few expert knowledge

Outline

- **The formal Neuron**

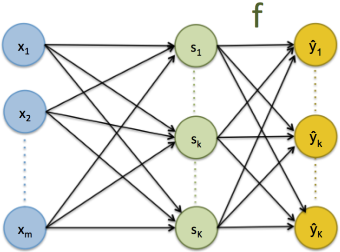


x_i: inputs
w_i, b: weights
f: activation function
y: output of the neuron

$$y = f(w^\top x + b)$$

Figure: The formal neuron – **Credits: R. Herault**

- <u>**Neural Networks:**</u> Stacking several formal neurons ⇒ **Perceptron**



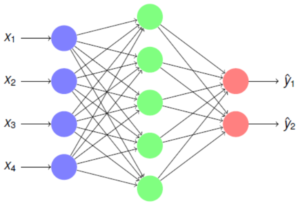
- **Soft-max Activation**:

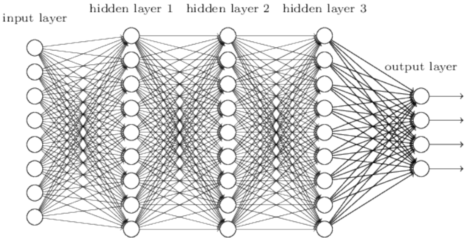
$$\hat{y}_k = f(s_k) = \frac{e^{s_k}}{\sum\limits_{k'=1}^{K} e^{s_{k'}}}$$

⇒ **Logistic Regression (LR) Model !**

Deep Neural Networks (DNN)

- **Multi-Layer Perceptron (MLP):** Stacking layers of neural networks
  - More complex and rich functions / Logistic Regression (LR)
  - **Neural network with one single hidden layer ⇒ universal approximator** [Cybenko, 1989]

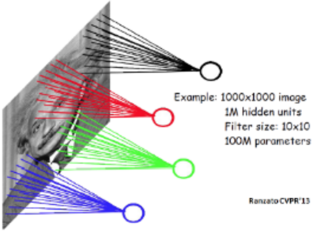


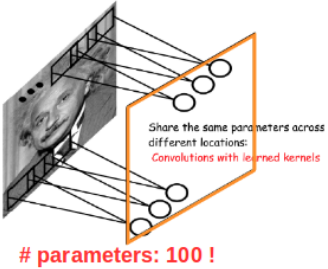


- **Basis of the "deep learning" field**
  - **Hidden layers: intermediate representations from data**
  - **Can be learned with Backpropagation algorithm [Lecun, 1985, Rumelhart et al., 1986]** (chain rule)

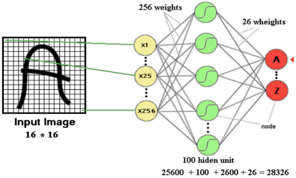
Convolutional Neural Networks (ConvNets)

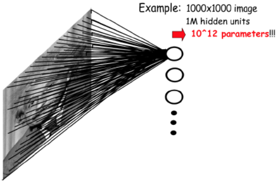
- **ConvNets:** sparse connectivity + shared weights





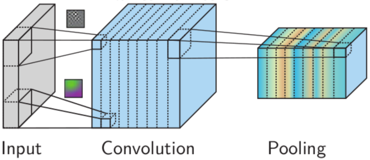
- **Local feature extraction ($\neq$ FCN)**
- Overcome parameter explosion for FCN on images

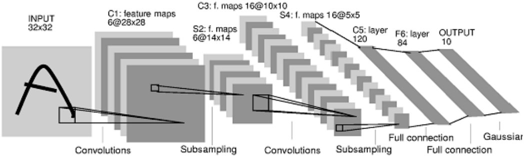




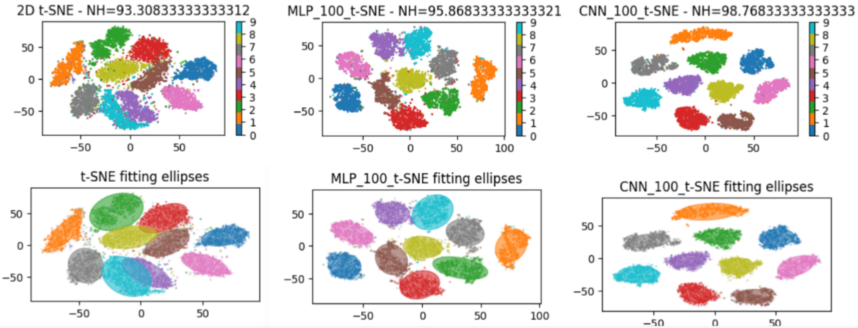
Convolutional Neural Networks (ConvNets)

- **Elementary block: Convolution + Non linearity (*e.g.* ReLU) + pooling**
  - **Stacking: deep ConvNets [LeCun et al., 1989]**



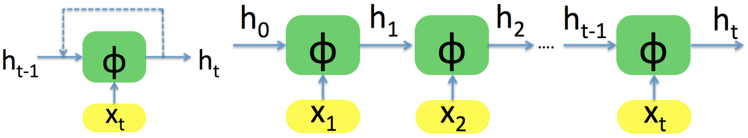


- **Parameters ↓, invariance ⇒ ↑ generalization & manifold disentangling!**



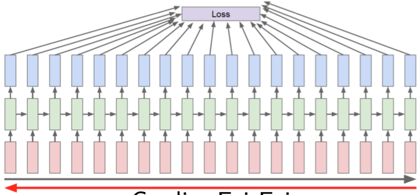
Recurrent Neural Networks (RNNs)

- **RNN Cell:** $\mathbf{h}_t = \phi(\mathbf{x}_t, \mathbf{h}_{t-1}) = f(\mathbf{U}\mathbf{x}_\mathbf{t} + \mathbf{W}\mathbf{h}_{\mathbf{t-1}} + \mathbf{b}_\mathbf{h})$ [Elman, 1990]
  - Loop, $\mathbf{h}_t$ depends on current $\mathbf{x}_t$ and previous state $\mathbf{h}_{t-1}$
    - $\mathbf{h}_t$: **network memory up to time t** $\Rightarrow$ **Sequence processing**
- Universal program [Siegelmann and Sontag, 1995] approximators



Folded RNN Unfolded RNN

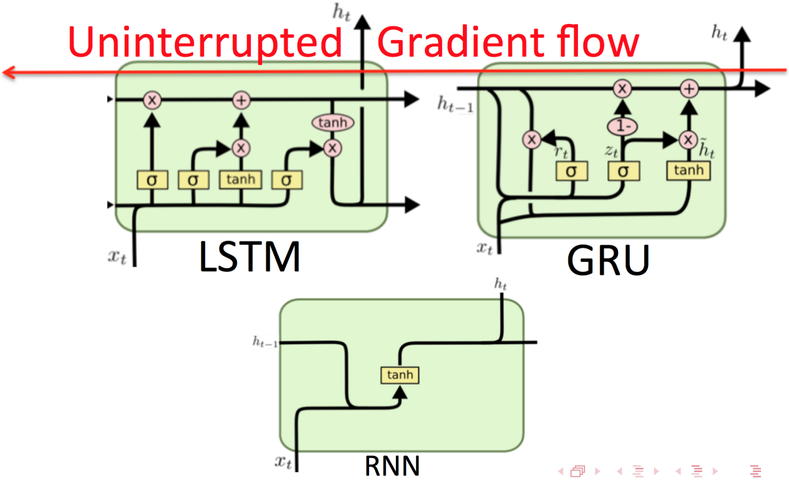
- Can be trained with Back-Propagation Through Time (BPTT)



Credit: Fei-Fei

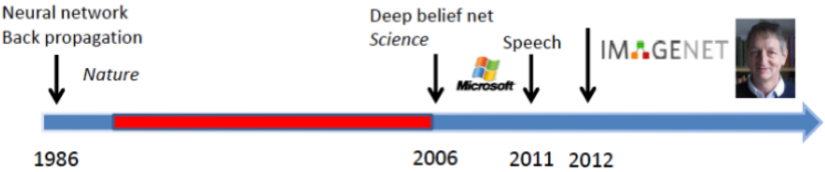
Recurrent Neural Networks (RNNs)

- **BUT Back-Propagation Through Time ⇒ vanishing gradients**
- **Specific architectures:**
  LSTM [Hochreiter and Schmidhuber, 1997], GRU [Cho et al., 2014]
  - LSTM: Cell gate ⇒ uninterrupted gradient flow



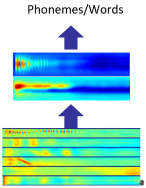
Deep Learning Succes since 2010

- 90's / 2000's: difficult to train large ConvNets / RNNs on big data



- Deep Learning renewal since 2010
  - 2011: Speech Recognition

Acoustic model	Recog \ WER	RT03S FSH	Hub5 SWB
Traditional features	1-pass −adapt	**27.4**	**23.6**
Deep Learning	1-pass −adapt	**18.5** (−33%)	**16.1** (−32%)



@Socher

Deep Learning Success since 2010

- **Deep Learning and ConvNet for Image Classification**
  - ImageNet ILSVRC Challenge (Stanford):
    - $1,200,000$ training images, $1,000$ classes, mono-label
    - Based on WordNet hierarchy (ontology)
  - Up to 2012, leading approaches: handcrafted features + shallow ML (SVM)
  - **ILSVRC'12: the deep revolution**
    $\Rightarrow$ **outstanding success of ConvNets [Krizhevsky et al., 2012]**

Rank	Name	Error rate	Description
1	**U. Toronto**	0.15315	Deep learning
2	U. Tokyo	0.26172	Hand-crafted features and learning models. Bottleneck.
3	U. Oxford	0.26979	
4	Xerox/INRIA	0.27058	

- **RNNs SOTA for many sequential decision making tasks:** speech, translation, text/music generation, times series, *etc*

Deep ConvNet success at ILSVRC'12

Two main practical reasons:

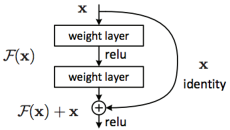
1. Huge number of labeled images (10^6 images)
    - Possible to train very large models without over-fitting
    - Larger models enables to learn rich (semantic) features hierarchies
2. GPU implementation for training
    - Relatively cheap and fast GPU
    - Training time reduced to 1-2 weeks (up to 50x speed up)

IMAGENET

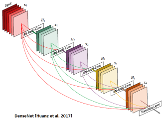
Current Trends in Deep Learning

Feature design $\Rightarrow$ network architecture design

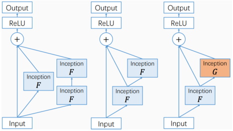
- **Improved training properties**, *e.g.* Res-Net or DenseNet ~ LSTM



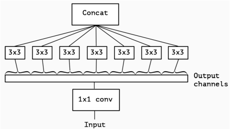
ResNet
[He et al., 2016]



DenseNet
[Huang et al., 2017]

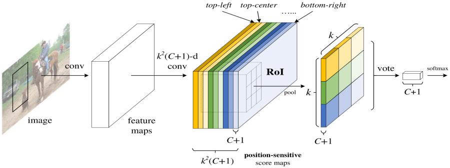


PolyNet
[Zhang et al., 2017]



Xception
[Chollet, 2017]

- **Combining blocks for specific tasks**, *e.g.* detection or ConvLSTM



[Dai et al., 2016]

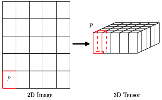


Figure 1: Transforming 2D image into 3D tensor

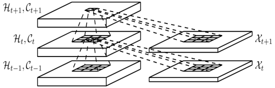
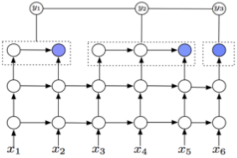


Figure 2: Inner structure of ConvLSTM

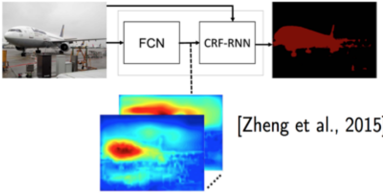
[Shi et al., 2015]

Training Models

- **Combining DL & structured prediction**, *e.g.* Conditional Random Fields (CRF)
  - Speech recognition (RNN+CRF), Semantic segmentation (ConvNets+CRF/RNN)

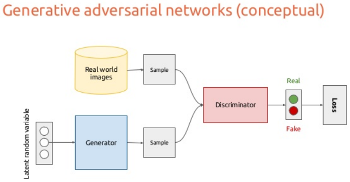


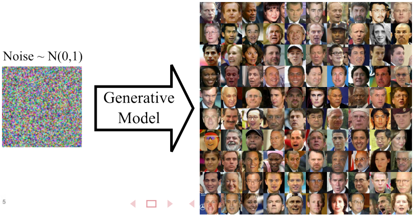
[Lu et al., 2016]



[Zheng et al., 2015]

- **Generative Adversarial Networks**: Game Theory (generator *vs* discriminator)
  - Adversarial cost used beyond generation for distribution matching

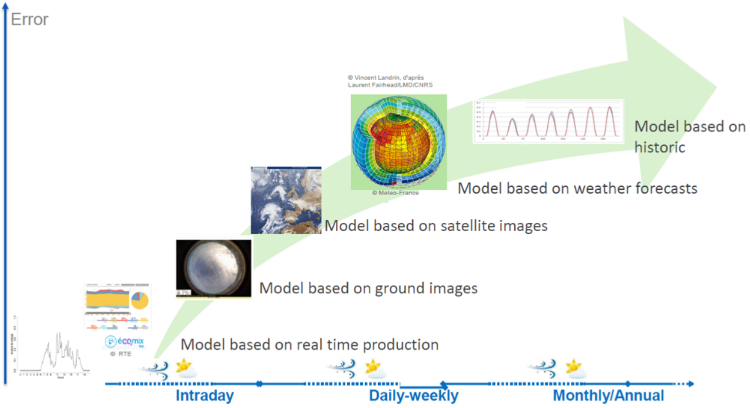




Outline

Context: PhotoVoltaic (PV) energy forecasting

Different data sources for different horizons

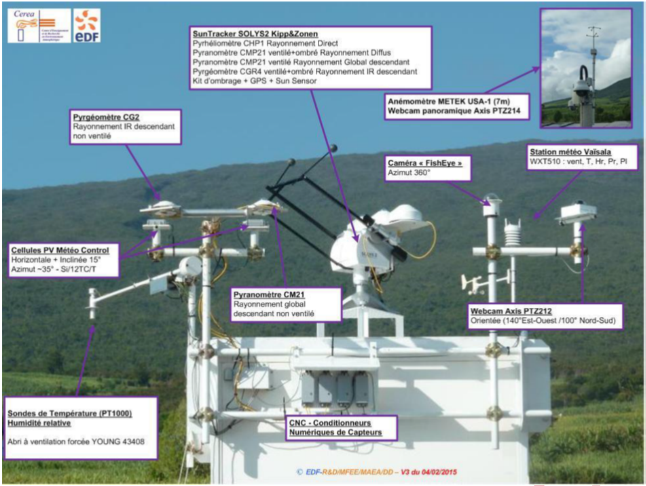


- **Ground based images:** very short term spatial & temporal horizons (0-20 min)
- **Application:** dynamic control of a hybrid system with PV, storage, diesel,...

Data

Meteorological campaign EDF R&D

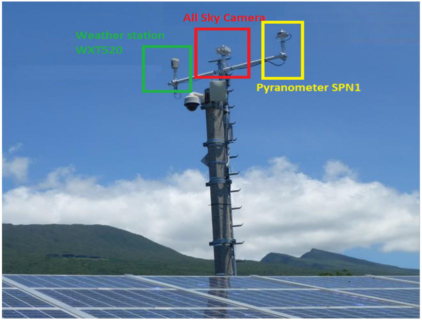
- EDF R&D experimental test site at La Reunion since 2012
- Many devices evaluated for solar resource assessment

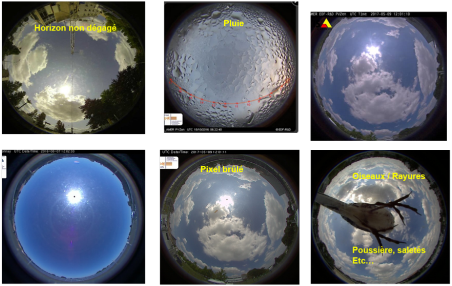


Data

Meteorological campaign EDF R&D

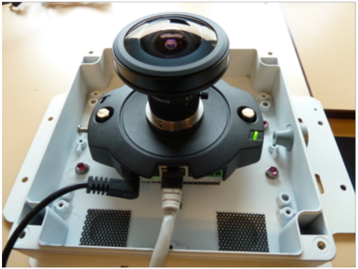
- **Choice:** ground images + pyranometer
- **Goal:** Can we use low-cost cameras instead of pyranometers to estimate current and future solar irradiations?





⇒ **More than 7 Millions images and corresponding irradiation measurements collected since 2010**

Fisheye camera: 180° hemispheric view of the sky, images every 10s

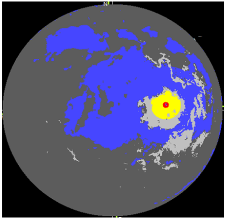



- **Pyranometer**: solar irradiance measurements every 10s
  - **GHI: Global Horizontal Irradiance**
  - DHI: Diffuse Horizontal Irradiance
  - DNI: Direct Normal Irradiance
- **Preprocessing:** irradiance values normalized by a clear sky model to remove seasonality $\Rightarrow$ $\boxed{\textbf{KGHI}}$

Baseline

Irradiance estimation module

1. **Image segmentation**: with handcrafted thresholds on Luminance and R-B



sun
circumsolar
sky
heavy clouds
light clouds

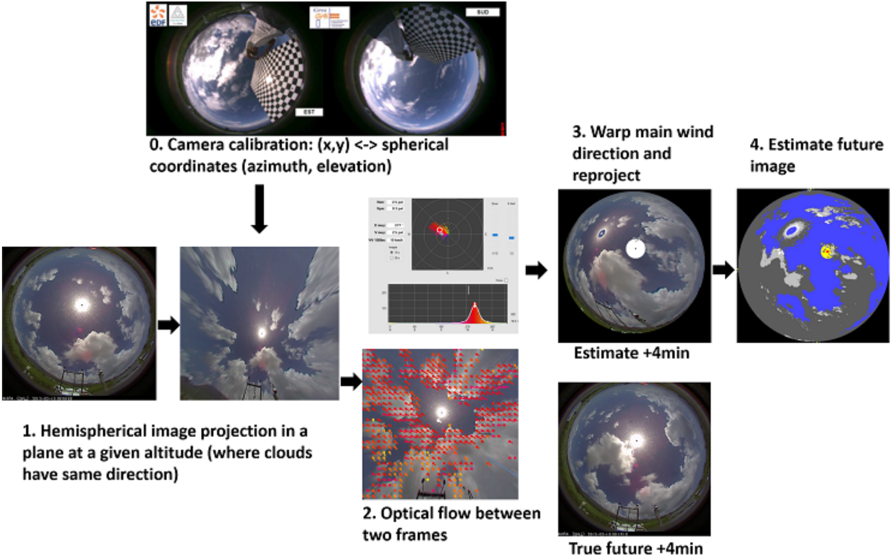
   - Each image described with feature $\mathbf{x_i} \in \mathbb{R}^5 \Leftrightarrow$ pixel class ratios
   - **Database**: $(\mathbf{x_i}, y_i)_{i=1:N}$ with $\mathbf{x_i}$ images and y_i corresponding KGHI

2. **Estimation**: kernel regression (Nadaraya-Watson model [Nadaraya, 1964]) for an unknown image x_0:

$$\widehat{y}(x_0) = \frac{C}{N} \sum_{i=1}^{N} e^{-\frac{||x_0 - x_i||^2}{2h^2}} y_i$$

Baseline

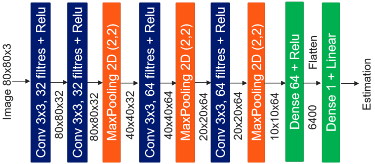
Forecasting module



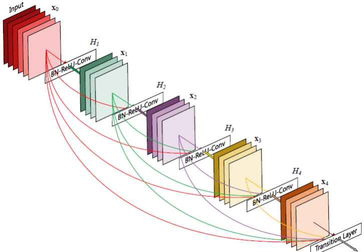
Estimating solar irradiance with ConvNets

Proposed neural network models

- **Small Convnet**: 475 000 parameters



- **Densenet model**: Densenet conv layers + dense regression layers
  201 layers: 18 Millions parameters



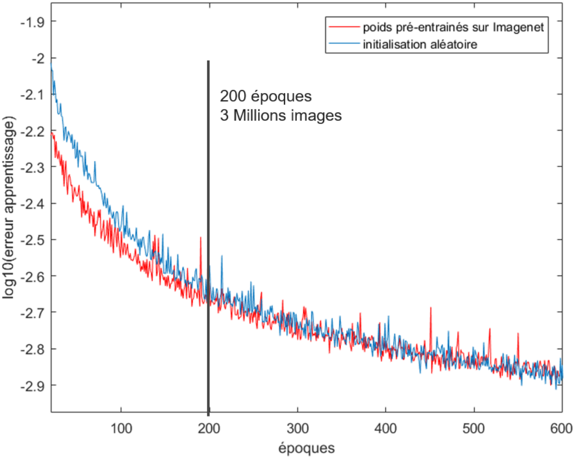
DenseNet [Huang et al. 2017]

- **Experimental setup:**
  - Training set: years 2012-2015 (4 190 064 images)
  - Test set: year 2016 (1 265 717 images)
- **Implementation:**
  - Python with Keras & Tensorflow backend
  - Adam optimizer
- **Training time:** - Nvidia Quadro P6000 (24 Go RAM)
  - 1 day for ConvNet, 6 days for DenseNet on a
- **Results for KGHI Estimation** (test set): large improvements of ConvNets wrt baseline

Model	Baseline	ConvNet	DenseNet
Mean Absolute Error (MAE)	0.1010	0.0448	**0.0197**
Normalized[1] MAE	14.9 %	6.59 %	**2.90 %**
Root Mean Square Error (RMSE)	0.1467	0.06992	**0.0328**
Normalized RMSE	21.6 %	10.3 %	**4.83 %**

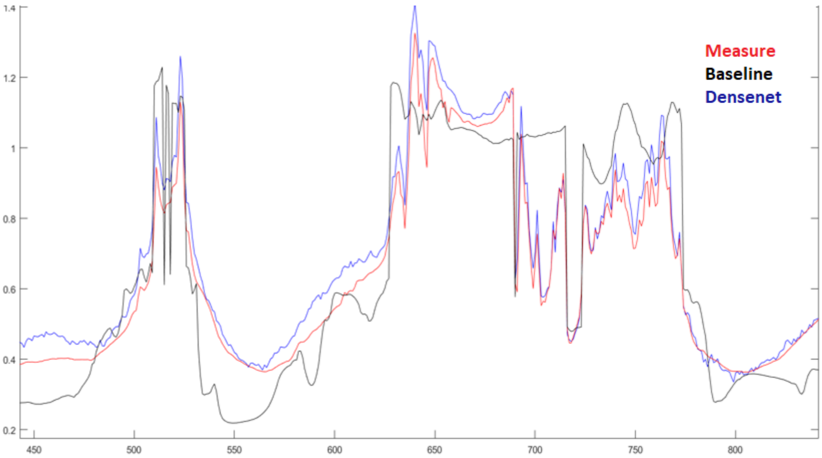
[1] by the mean KGHI value over the training set

Learning from scratch possible (ImageNet pre-training not necessary)



Estimating solar irradiance with ConvNets

Results on a particular day



Estimating solar irradiance with ConvNets

t-SNE vizualization

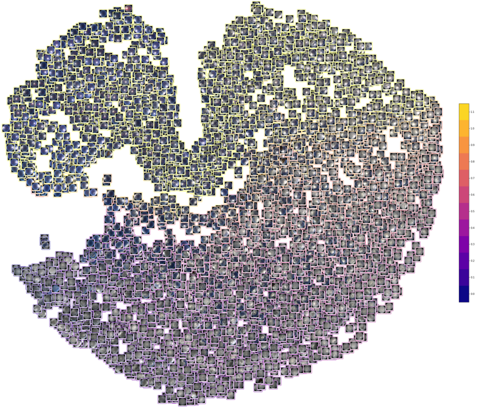
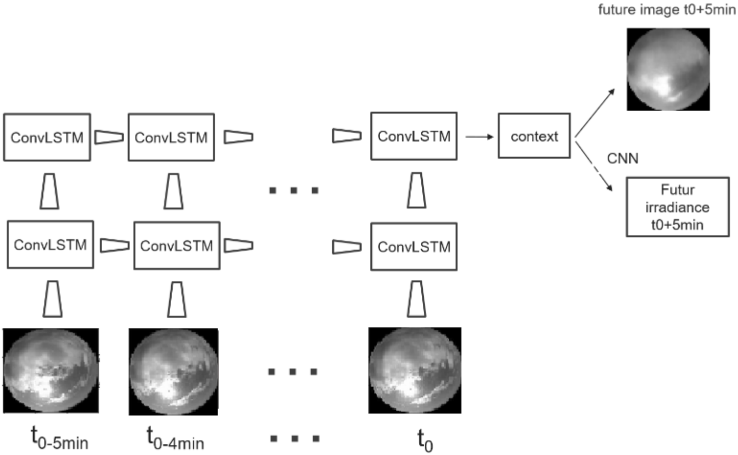


Figure: Clustering on Densenet features. Upper left: clear sky, upper right: cloudy, bottom: very cloudy, rainy

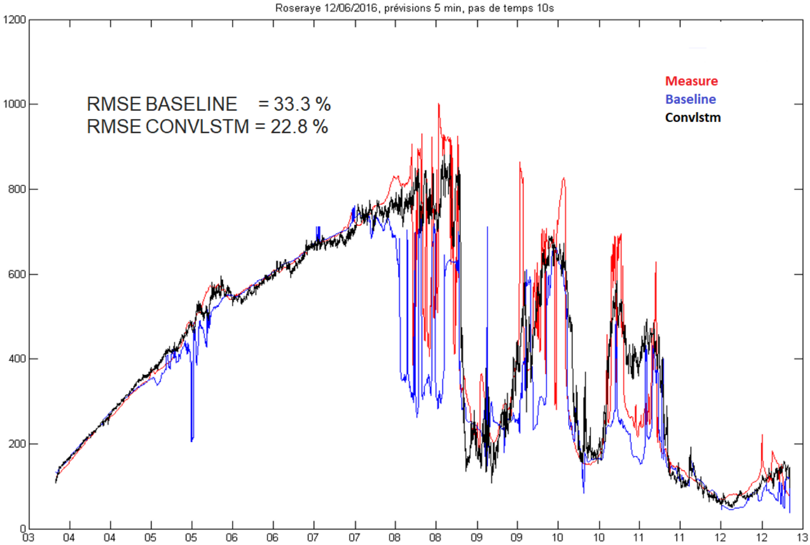
Ongoing work: irradiance forecasting

- **First proposed model:** light for computational purposes
- Input sequence: 10 grayscale 60x60 images every 30s
- Predict future image and irradiance at 5min
- Stacked ConvLSTM layers as spatiotemporal feature extractor



Ongoing work: irradiance forecasting

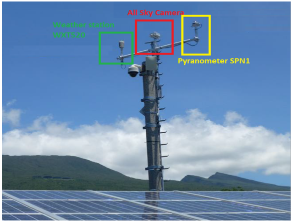
Preliminary results

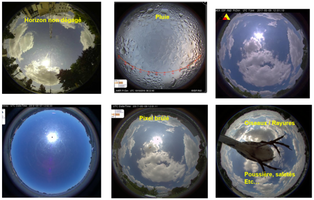


Outline

Conclusion & Perspectives

- **Effective deep ConvNet solutions for irradiance predictions on static images**
  - **Favorable context:** huge volume of annotated data





- **Promising first results for future irradiance forecasting**
  - Short-term: contribution of temporal information to irrandiance & forecasts
  - Longer term: improve forecasting models and training methodologies

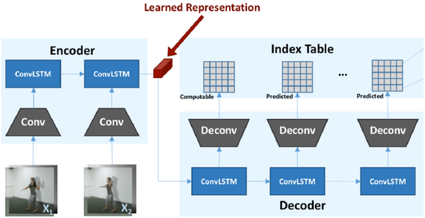
Deep learning for video prediction

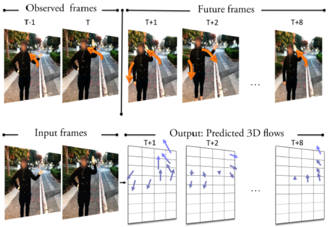
- Direct RGB future image generation still challenging for large and complex natural images, predictions become blurry [Srivastava et al., 2015]
- To mitigate this: learn geometric transforms between images [Finn et al., 2016], use an adversarial loss instead of L2 [Mathieu et al., 2015]

[Mathieu et al. 2015]

Direct irradiance prediction with deep learning without future image prediction

- Predict latent features in encoder-decoder RNN architectures [Luo et al., 2017]





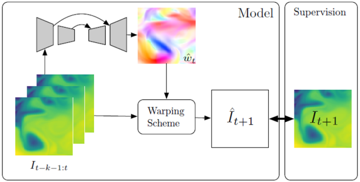
- Predict future image segmentations [Luc et al., 2018]

Forecasting future irradiances

Direct irradiance prediction

Prediction with physical knowledge

- Introduce a priori physical information (advection diffusion PDE) [de Bezenac et al., 2018]



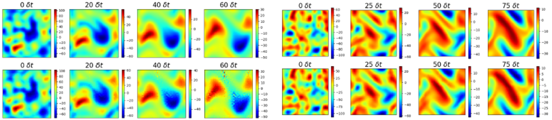
Advection- diffusion equation

$$\frac{\partial I}{\partial t} + (w.\nabla)I = D\nabla^2 I$$

Warping scheme

$$\hat{I}_{t+1}(x) = \sum_{y\in\Omega} k(\, x - \hat{w}(x),\, y)\, I_t(y)$$

- Approximate differential equation solutions with neural nets [Long et al., 2017, Chen et al., 2018a]



[Long et al. 2015]

(a) 5 × 5 filters. (b) 7 × 7 filters.

Figure 5: Images of the true dynamics and the predicted dynamics. The first row shows the images of the true dynamics. The second row shows the images of the predicted dynamics using the PDE-Net

Choice of loss function

How to penalize delays on the predicted irradiance time series ?

Goal: forecast irradiance ramps on time

Classical loss functions (MAE, RMSE) ill adapted to distinguish absolute value errors from temporal distorsion errors. Possible solutions:

- signal gradient loss [Mathieu et al., 2015]
- loss based on Dynamic Time Warping [Cuturi and Blondel, 2017],...

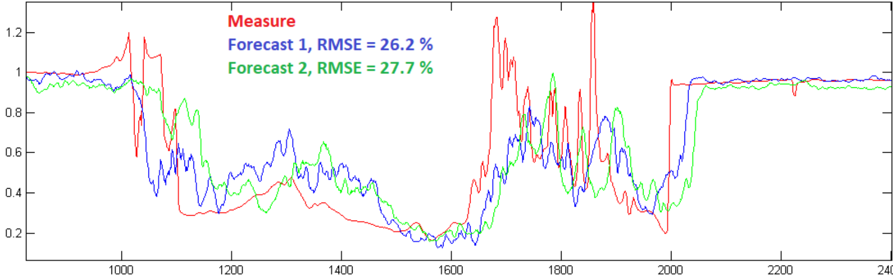
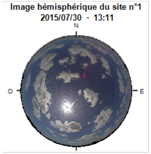


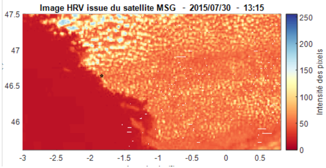
Figure: 2 forecasts with similar RMSE but different anticipating skills. Forecast 2 model seems always late

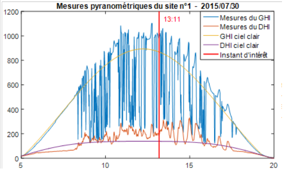
Multi-modal fusion

Ground images, satellite images, numerical weather forecasts

- Improve forecasting temporal and spatial scales by fusing different data sources:
  - PV production measurements
  - ground based camera
  - satellite image
  - numerical weather forecasts







- Use a network of multiple cheap cameras on a territory to anticipate global phenomena

Thank you for your attention!

Questions?

References I

[Chen et al., 2018a] Chen, T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018a).
Neural ordinary differential equations.
NIPS 2018.

[Chen et al., 2018b] Chen, W., Alexis, M., Pierre, T., Justin, S., Nicolas, L., Guillaume, E., Ralph, F., Douglas, V., and Bertrand, C. (2018b).
Labeled sar imagery dataset of ten geophysical phenomena from sentinel-1 wave mode (tengeop-sarwv).
SEANOE.

[Cho et al., 2014] Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014).
Learning phrase representations using rnn encoder-decoder for statistical machine translation.
cite arxiv:1406.1078Comment: EMNLP 2014.

[Chollet, 2017] Chollet, F. (2017).
Xception: Deep learning with depthwise separable convolutions.
In *CVPR*, pages 1800–1807. IEEE Computer Society.

[Cuturi and Blondel, 2017] Cuturi, M. and Blondel, M. (2017).
Soft-dtw: a differentiable loss function for time-series.
arXiv preprint arXiv:1703.01541.

[Cybenko, 1989] Cybenko, G. (1989).
Approximation by superpositions of a sigmoidal function.
Mathematics of control, signals and systems, 2(4):303–314.

[Dai et al., 2016] Dai, J., Li, Y., He, K., and Sun, J. (2016).
R-fcn: Object detection via region-based fully convolutional networks.
In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 379–387. Curran Associates, Inc.

[de Bezenac et al., 2018] de Bezenac, E., Pajot, A., and Gallinari, P. (2018).
Deep learning for physical processes: Incorporating prior scientific knowledge.
ICLR.

References II

[Elman, 1990] Elman, J. L. (1990).
Finding structure in time.
COGNITIVE SCIENCE, 14(2):179–211.

[Finn et al., 2016] Finn, C., Goodfellow, I., and Levine, S. (2016).
Unsupervised learning for physical interaction through video prediction.
In *Advances in neural information processing systems*, pages 64–72.

[He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016).
Deep residual learning for image recognition.
In *CVPR*.

[Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997).
Long short-term memory.
Neural Comput., 9(8):1735–1780.

[Huang et al., 2017] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017).
Densely connected convolutional networks.
CVPR 2017.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).
Imagenet classification with deep convolutional neural networks.
In *Advances in neural information processing systems*, pages 1097–1105.

[Lecun, 1985] Lecun, Y. (1985).
Une procedure d'apprentissage pour reseau a seuil asymmetrique (A learning scheme for asymmetric threshold networks), pages 599–604.

[LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989).
Backpropagation applied to handwritten zip code recognition.
Neural computation, 1(4):541–551.

References III

[Lguensat et al., 2018] Lguensat, R., Sun, M., Fablet, R., Tandeo, P., Mason, E., and Chen, G. (2018).
Eddynet: A deep neural network for pixel-wise classification of oceanic eddies.
In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1764–1767. IEEE.

[Long et al., 2017] Long, Z., Lu, Y., Ma, X., and Dong, B. (2017).
Pde-net: Learning pdes from data.
arXiv preprint arXiv:1710.09668.

[Luc et al., 2018] Luc, P., Couprie, C., Lecun, Y., and Verbeek, J. (2018).
Predicting future instance segmentations by forecasting convolutional features.
arXiv preprint arXiv:1803.11496.

[Luo et al., 2017] Luo, Z., Peng, B., Huang, D.-A., Alahi, A., and Fei-Fei, L. (2017).
Unsupervised learning of long-term motion dynamics for videos.
arXiv preprint arXiv:1701.01821, 2.

[Mathieu et al., 2015] Mathieu, M., Couprie, C., and LeCun, Y. (2015).
Deep multi-scale video prediction beyond mean square error.
arXiv preprint arXiv:1511.05440.

[Nadaraya, 1964] Nadaraya, E. A. (1964).
On estimating regression.
Theory of Probability & Its Applications, 9(1):141–142.

[Racah et al., 2017] Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, M., and Pal, C. (2017).
Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events.
In *Advances in Neural Information Processing Systems*, pages 3402–3413.

[Rumelhart et al., 1986] Rumelhart, D., Hinton, G., and Williams, R. (1986).
Learning representations by back-propagating errors.
Nature, 323:533–536.

References IV

[Shi et al., 2015] Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c. (2015).
Convolutional lstm network: A machine learning approach for precipitation nowcasting.
In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, pages 802–810, Cambridge, MA, USA. MIT Press.

[Siegelmann and Sontag, 1995] Siegelmann, H. and Sontag, E. (1995).
On the computational power of neural nets.
J. Comput. Syst. Sci., 50(1):132–150.

[Srivastava et al., 2015] Srivastava, N., Mansimov, E., and Salakhudinov, R. (2015).
Unsupervised learning of video representations using lstms.
In *International conference on machine learning*, pages 843–852.

[Zhang et al., 2017] Zhang, X., Li, Z., Loy, C. C., and Lin, D. (2017).
Polynet: A pursuit of structural diversity in very deep networks.
In *CVPR*, pages 3900–3908. IEEE Computer Society.