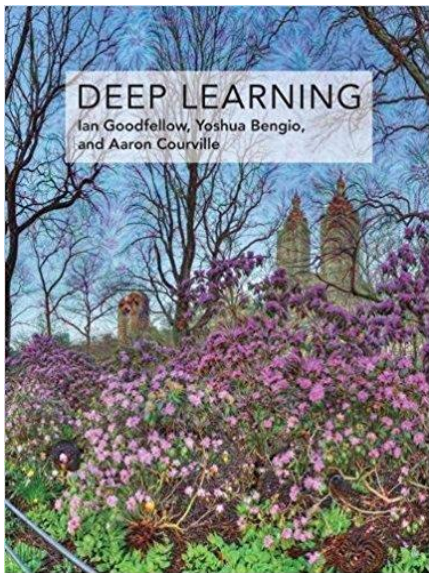# 2017-2018 Recap

DeepLearningPlayground

OBJECTIVE
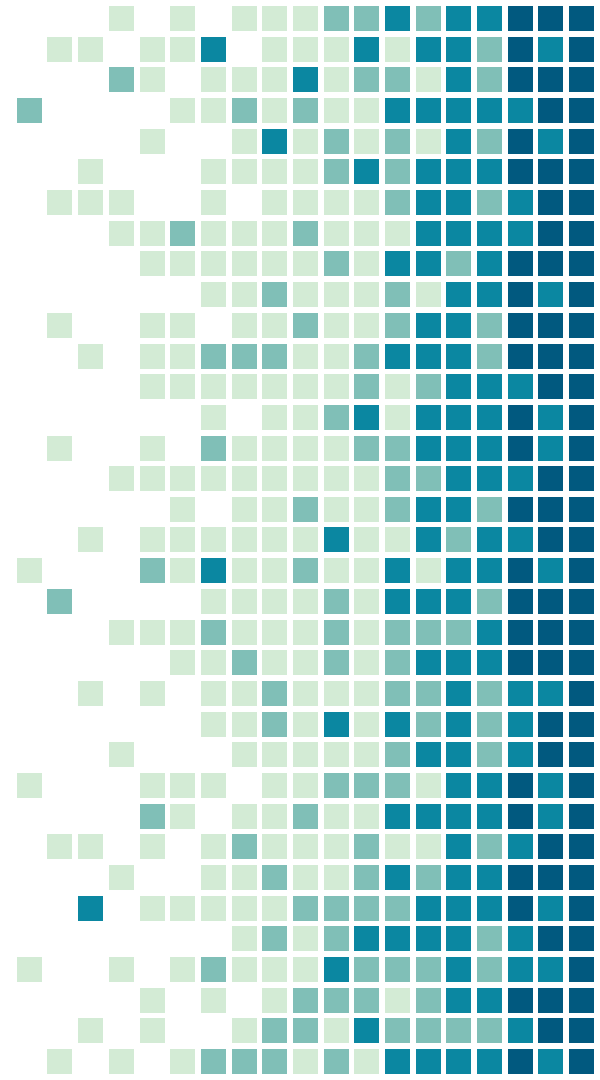
Put theorical concepts into practice !

Tensorflow, Jupyter, Tensorboard...
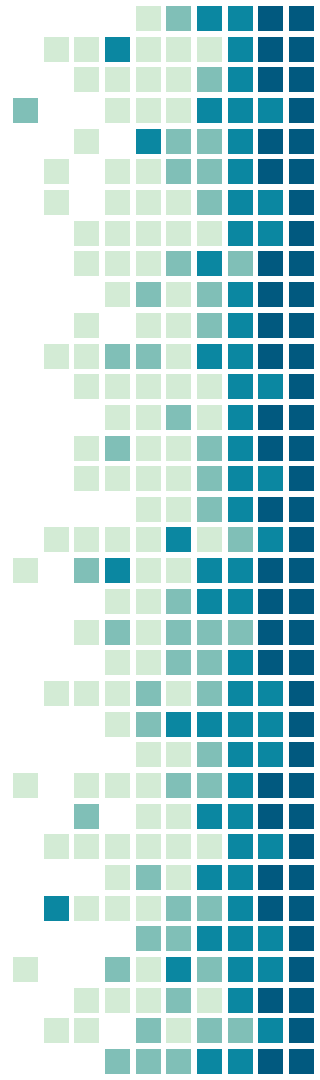
1.

CHAPTERS 6-7-8 : FeedForward & Optimization

1 Sessions

CONCEPTS

- <u>Optimizers</u> : SGD, Adam, RMSProp
- <u>Regularization</u> : L2, Dropout
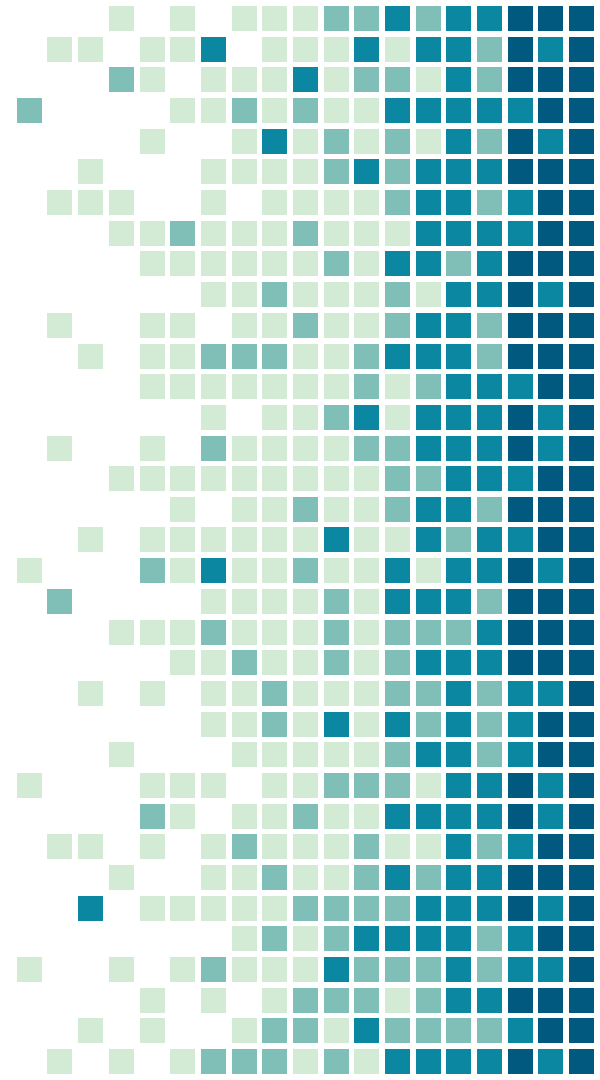- Early Stopping

- *(Gradient Clipping, Batch Normalization...)*
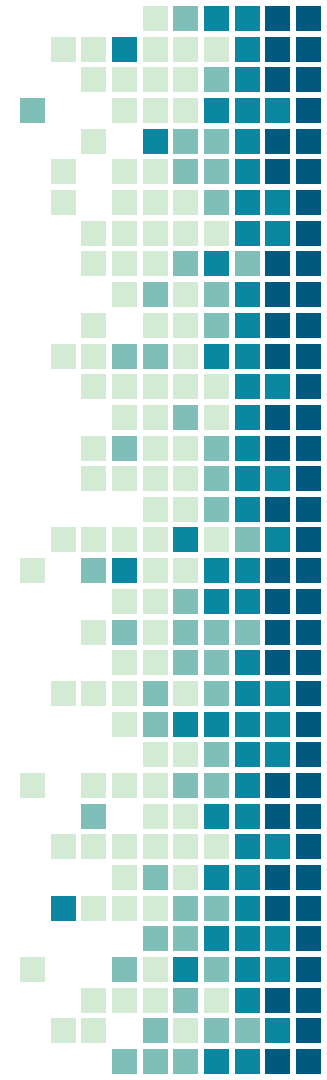
2.
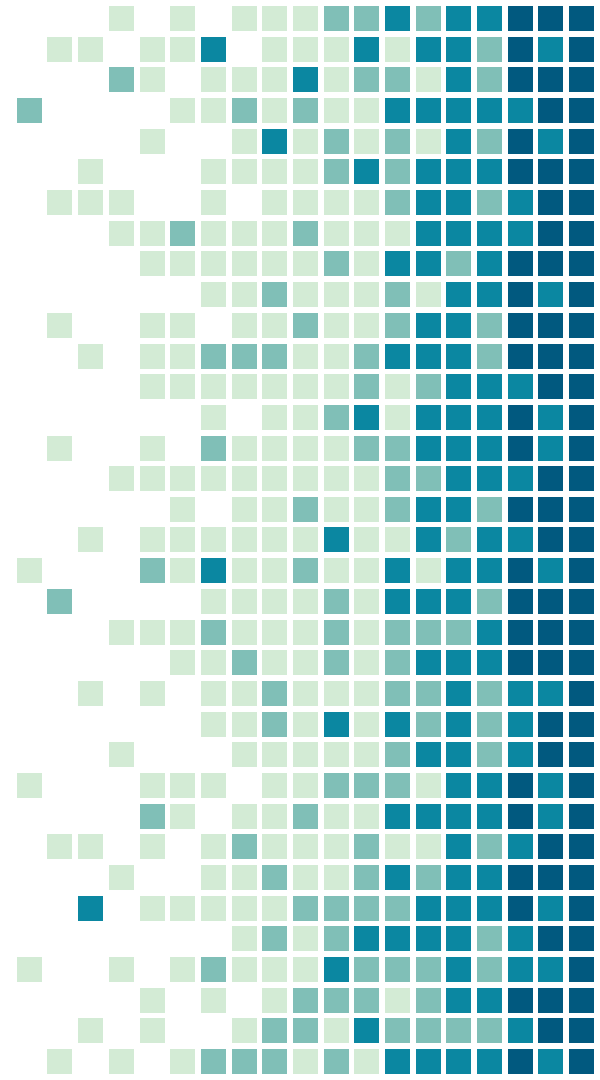# CHAPTER 9 :
CNNs

2 Sessions

CONCEPTS

- Convolutional Layers
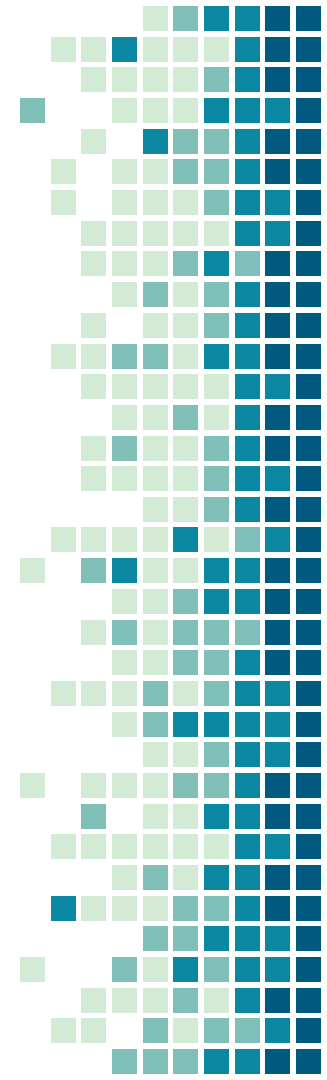- Max Pooling
- Zero Padding
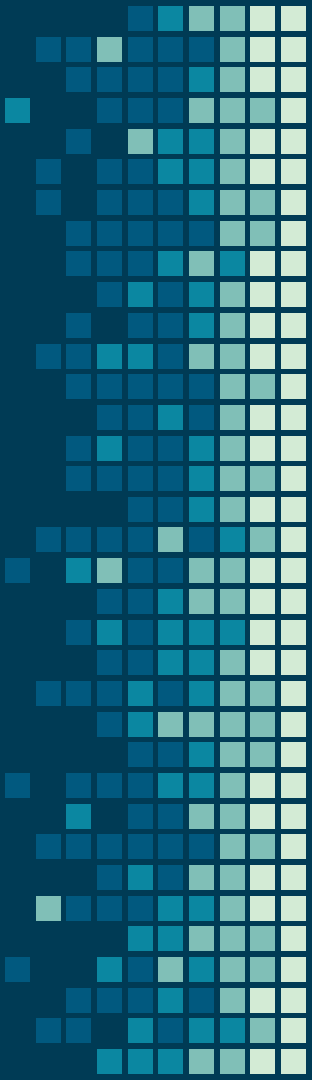
3.
CHAPTER 10 : RNNs

5 Sessions

CONCEPTS

- Vanilla RNNs
- Encoders-Decoders (Seq2Seq)
- LSTM

- *(Other GRUs)*
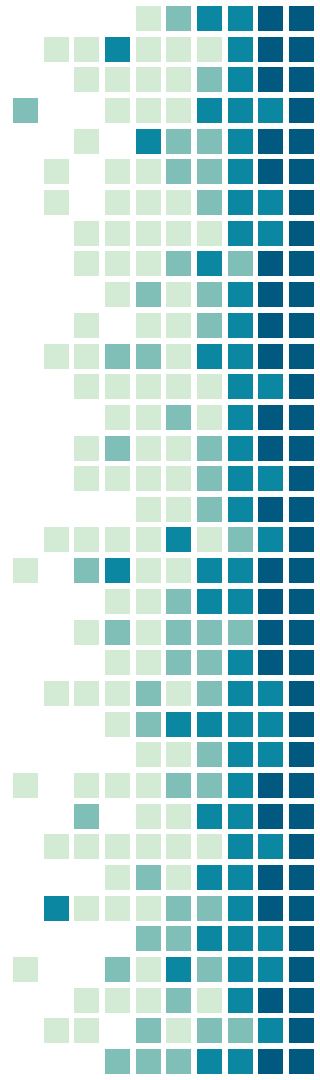
Neural Machine Translation

DATASET *(BEFORE PROCESSING)*

- European Parliament proceedings (1996)
- 11 languages
- French to English

2,007,723 sentences

51,388,643 French words
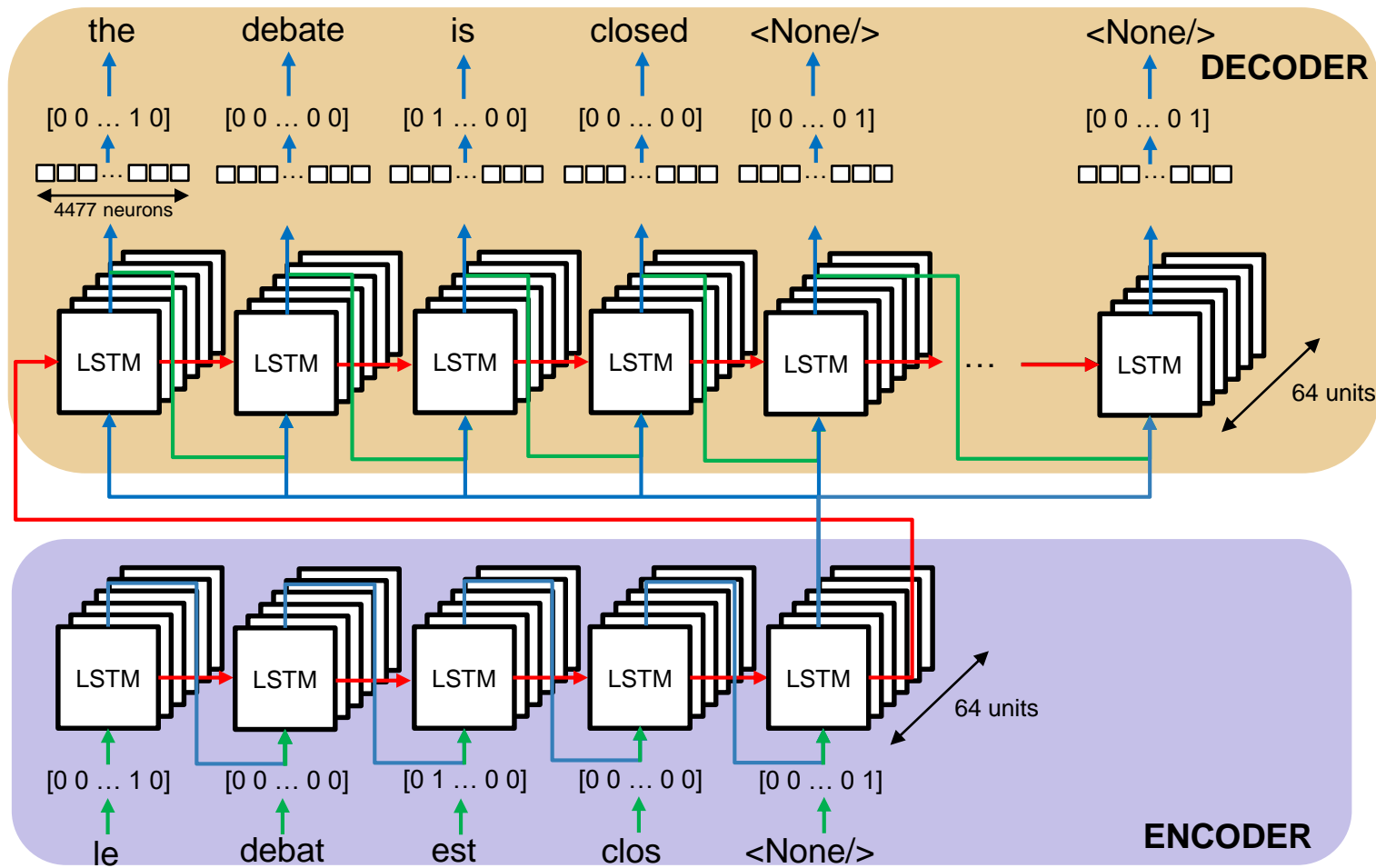
50,196,035 English words

DATASET *(AFTER PROCESSING)*

- Replaced barely used words by token : **<5**
- Kept only short sentences : **5 words max in output**
- Removed anomalies : **len(fr) > 3*len(en) || len = 0**
- Reduced to **10 000 sentences**

10,000 sentences (90% Train/Test)

5,636 French words

4,477 English words

ARCHITECTURE

the   debate   is   closed   <None/>   <None/>

DECODER

[0 0 … 1 0]   [0 0 … 0 0]   [0 1 … 0 0]   [0 0 … 0 0]   [0 0 … 0 1]   [0 0 … 0 1]

4477 neurons

LSTM LSTM LSTM LSTM LSTM … LSTM

64 units

[0 0 … 1 0]   [0 0 … 0 0]   [0 1 … 0 0]   [0 0 … 0 0]   [0 0 … 0 1]

LSTM LSTM LSTM LSTM LSTM
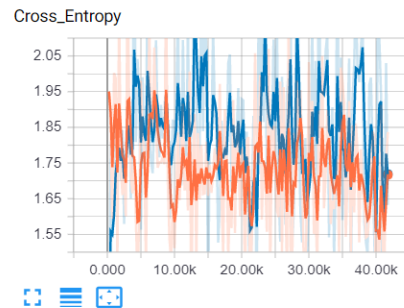
64 units

ENCODER

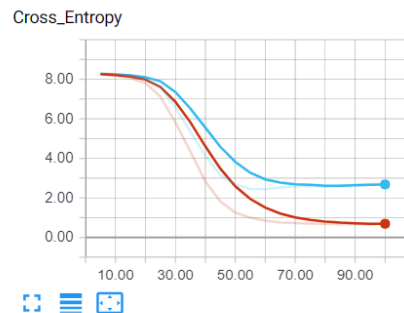le   debat   est   clos   <None/>

12

TRAINING

1) First training :

   Not learning !

2) Implementation validation :

   Reduced TrainSet to **5 samples**

   Check if **Overfitting**

TRAINING

3) Diagnostic :

- Sentences with empty input/output

- Almost whole output padded => learns to just pad

```
Input : [['nous', 'devons', 'etre', 'flexibles', None]]
Output : [['we', 'must', 'be', 'flexible', None, None, None, None, None, None, None, None, None, None, None, None, None, None,
None]]
Predicted Output : [[None, None, None, None, None, None, None, None, None, None, None, None, None, None, None, None, None, Non
e, None]]
```
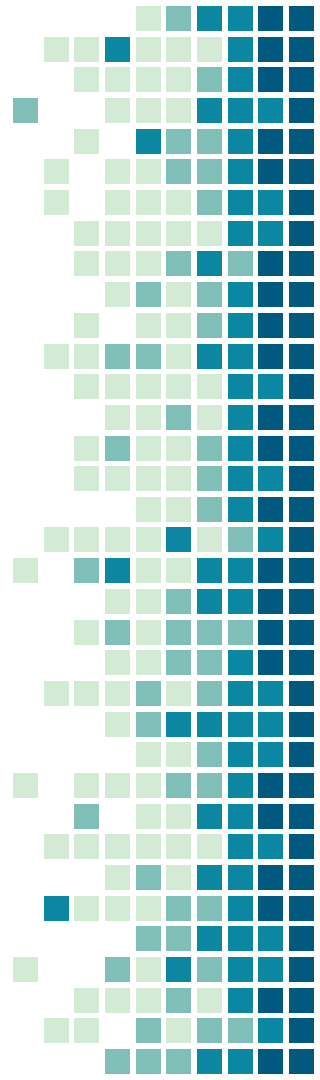
4) Corrections :

Remove anomalies

Mask padding in gradient calculation ?

WHAT'S NEXT ?

- Attention Mecanism
- Use Embeddings instead of OneHotEncoding
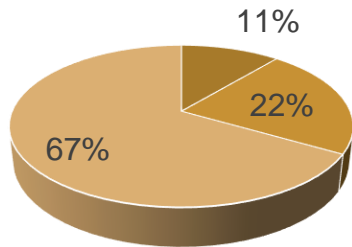- Improve the latent space

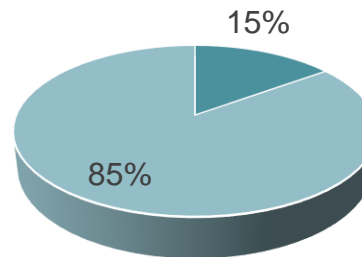# Conclusion

Chapters repartition 🕐

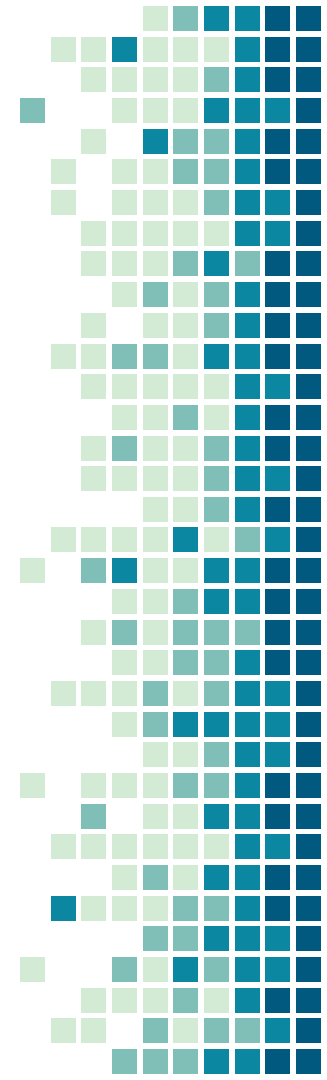10 Sessions

■ Chapters 6,7,8 ■ Chapter 9 ■ Chapter 10 ■

11%

22%

67%

Read / Dev 📅
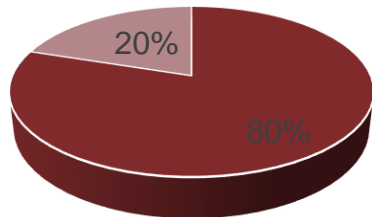
■ Read ■ Dev

15%

85%

Real time working ☕

■ Working ■ Other

20%

80%

THANKS!

Any questions?

You can find me at:

@ClementRomac

Github/ClementRomac