ARIADNEXT
WE DO IDENTITY.

PyraD-DCNN: a Fully Convolutional Neural Network to Replace BLSTM in Offline Text Recognition Systems

Jonathan Jouanne Quentin Dauchy
Ahmad Montaser Awal

11th of January 2021

CADL2020

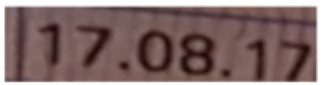
Optical Character Recognition

- Transcribe handwritten or printed text line images into strings
  - Pure 2D-signal processing such as **image semantic segmentation**
  - **Sequence processing** such as language modelisation

ARIADNEXT
WE DO IDENTITY.

Optical Character Recognition

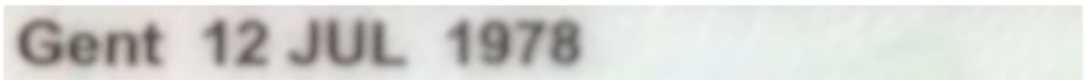
- Transcribe handwritten or printed text line images into strings
  - Pure 2D-signal processing such as **image semantic segmentation**
  - **Sequence processing** such as language modelisation



(a) Slanted delivery date

(b) Forname with polish diacritics

(c) Spouse name with strong background patterns



(d) Blurry place and date of birth

Introduction

- Hybrid neural networks provide state-of-the-art performance, stacking
  - Optical model: mainly **CNN layers** handling 2D information for graphical description
  - Cognitive model: mainly **RNN layers** handling sequential information
  - Interpretation by a Connectionist Temporal Classification heuristic layer

Introduction

ARIADNEXT
WE DO IDENTITY.

- Hybrid neural networks provide state-of-the-art performance, stacking
  - Optical model: mainly **CNN layers** handling 2D information for graphical description
  - Cognitive model: mainly **RNN layers** handling sequential information
  - Interpretation by a Connectionist Temporal Classification heuristic layer

But ..

- Inappropriate for embedded systems (ex. Smartphones) as they are time and storage consuming

Introduction

ARIADNEXT
WE DO IDENTITY.

- Hybrid neural networks provide state-of-the-art performance, stacking
  - Optical model: mainly **CNN layers** handling 2D information for graphical description
  - Cognitive model: mainly **RNN layers** handling sequential information
  - Interpretation by a Connectionist Temporal Classification heuristic layer

But ..

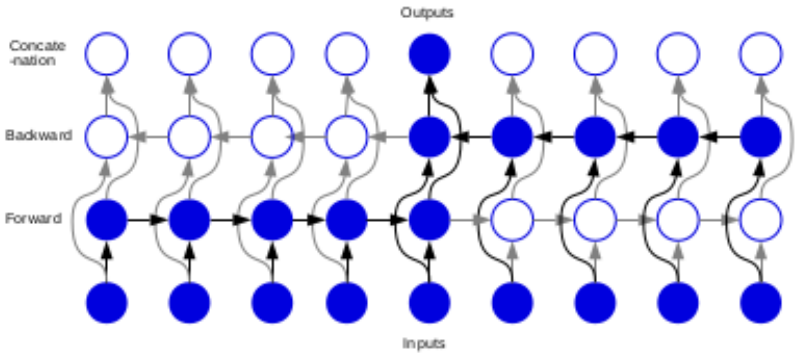
- Inappropriate for embedded systems (ex. Smartphones) as they are time and storage consuming

More recent works

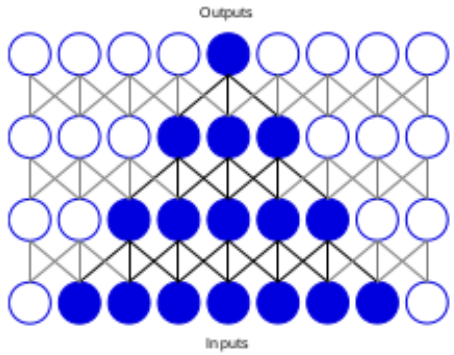
- more use of dilated convolution neural networks (DCNN)

Design principals

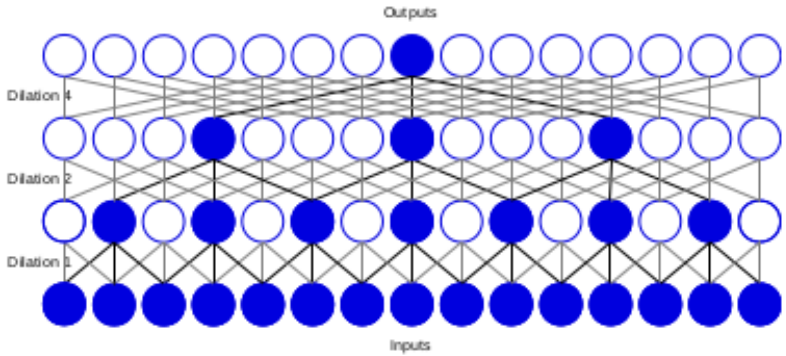
Dilated Convolutions



(a) Bi-directional LSTM



(b) Stack of three vanilla convolutions

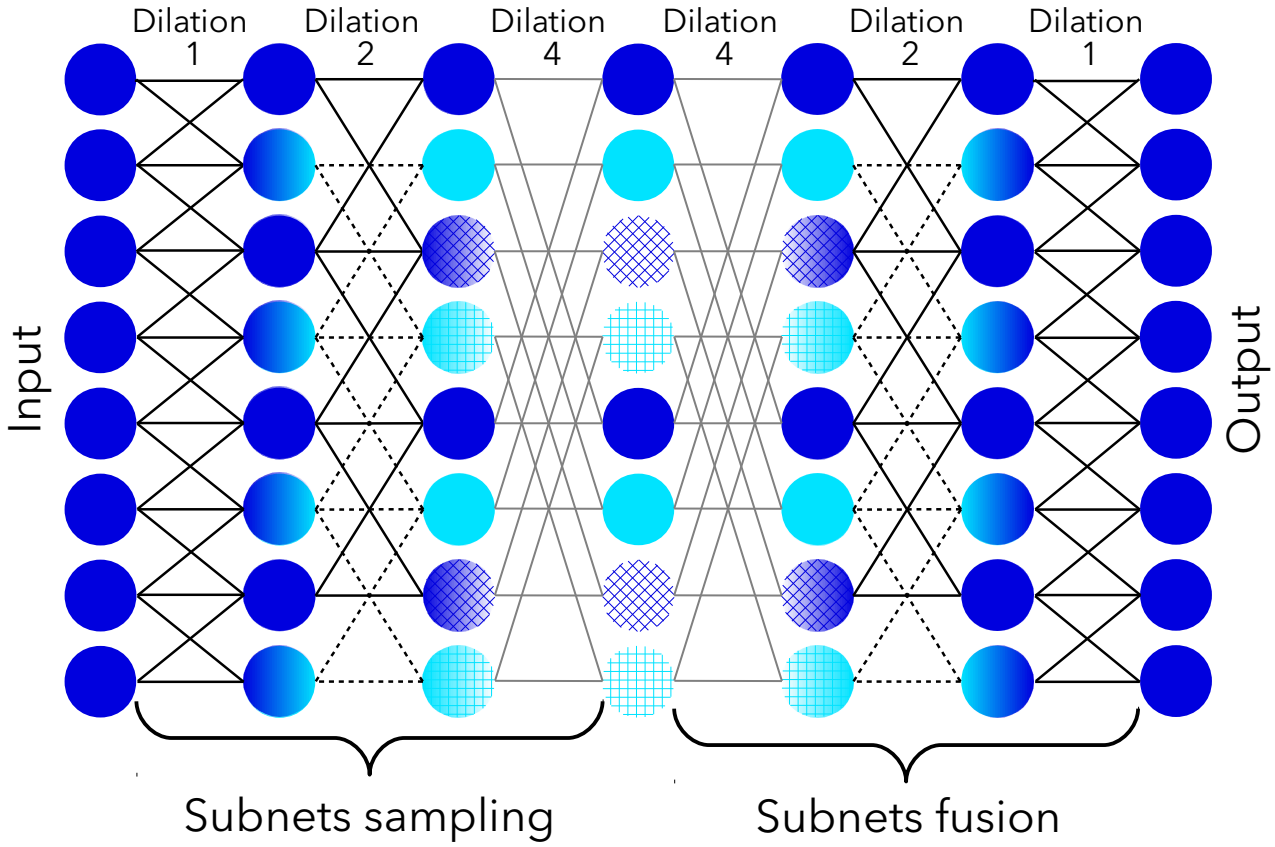


(c) Stack of dilated convolutions

Design principals

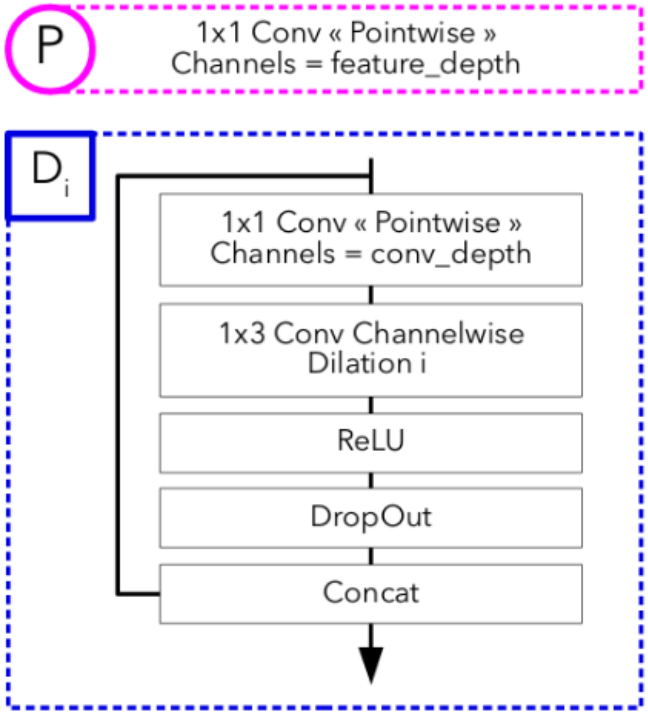
ARIADNEXT
WE DO IDENTITY.

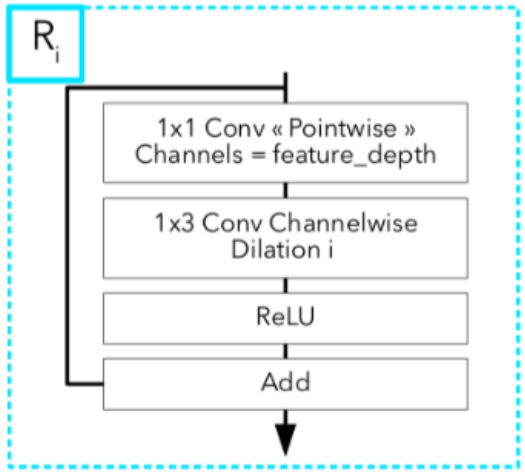
Multi-Resolution Pyramid with Dilated Networks

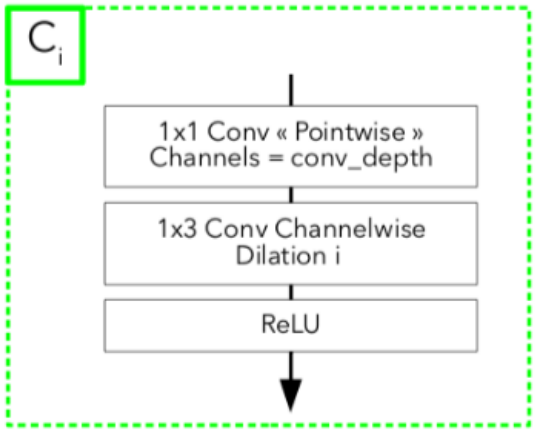


Design principals

Skip Connections and Bottlenecks







Overall architecture

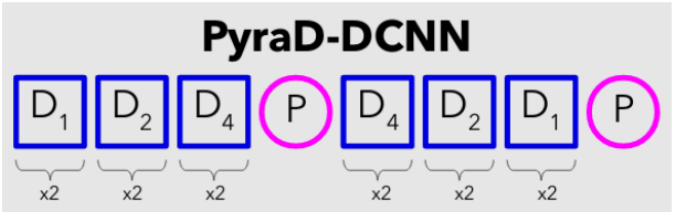
ARIADNEXT
WE DO IDENTITY.

- Common optical feature extractor
- PyraD-DCNN Sequence Model (Multi-resolution pyramid)
  - Three dilation stages
  - Two Dilated Convolution Blocks (DCB)
  - Connectionist Temporal Classification as the last layer

Overall architecture

ARIADNEXT
WE DO IDENTITY.

- Common optical feature extractor
- PyraD-DCNN Sequence Model (Multi-resolution pyramid)
  - Three dilation stages
  - Two Dilated Convolution Blocks (DCB)
  - Connectionist Temporal Classification as the last layer



Ablation study

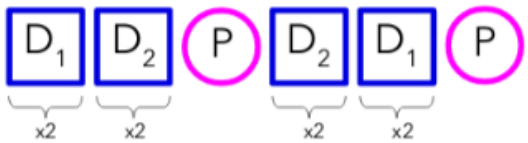
Families of derived architectures for the sequence modeling part based on

- A BLSTM network as Benchmark
- Pyramid height (shallower to deeper networks)
- Dilation Factor (vs. a flat model)
- Dilated convolutions stacking strategy (in terms of resolution)
- Bottleneck Pointwise Convolutions' positions and amounts
- Skip connections presence

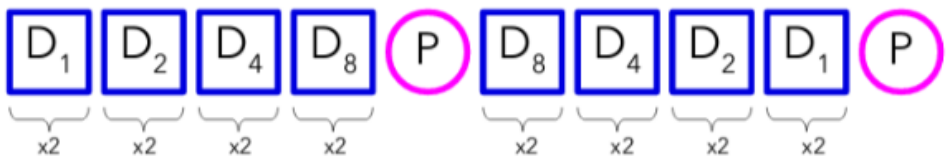
Ablation study

ARIADNEXT
WE DO IDENTITY.

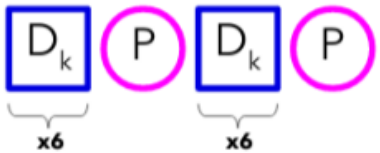
Shallow PyraD-DCNN



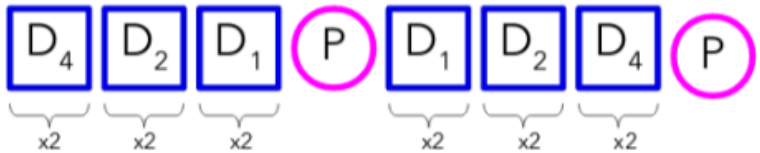
Deep PyraD-DCNN



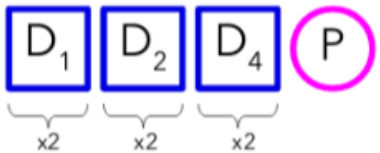
k-Flat-DCNN



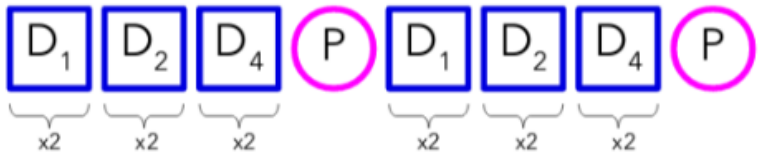
Reverse PyraD-DCNN



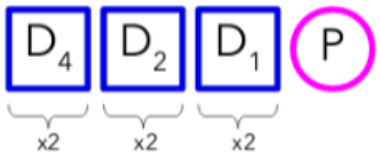
Up-DCNN



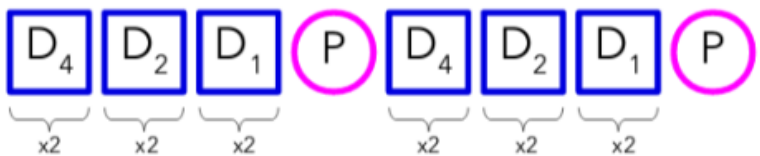
UpUp-DCNN



Down-DCNN

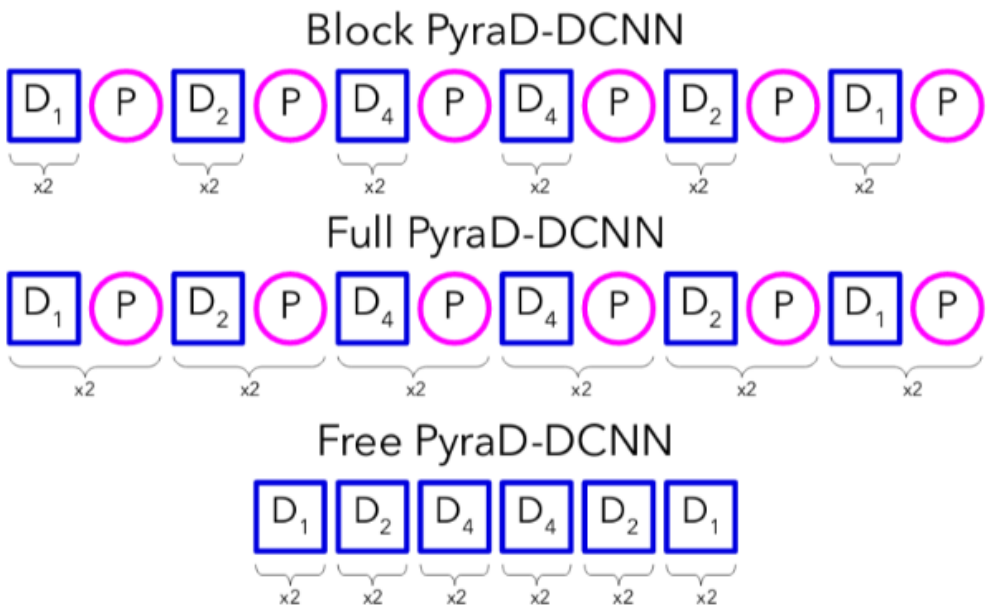


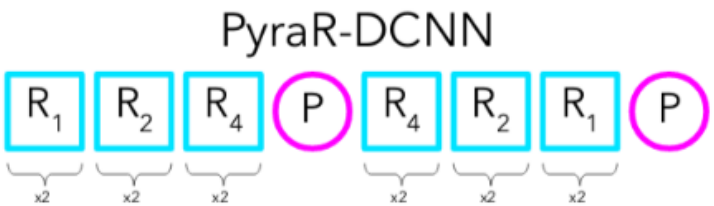
DownDown-DCNN

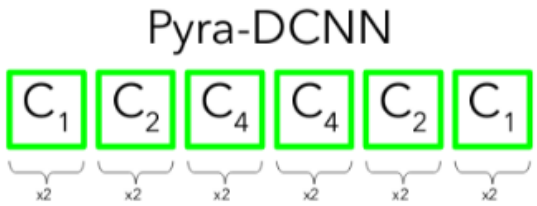


Ablation study

ARIADNEXT
WE DO IDENTITY.







Experimentation & Results

ARIADNEXT
WE DO IDENTITY.

- **Data-set**: text field images from various European identity documents
  - Small-ID: Totaling 40 000 samples (50 characters)
  - Big-ID: Totaling 600 000 samples (150 characters)
- Data split: 60% for training, 20% for validation and the last 20% for evaluation. No data augmentation has been used in this work
- Both perfect match rate (accuracy) and character error rate (CER) are considered

Experimentation & Results

ARIADNEXT
WE DO IDENTITY.

Observations upon experiments on SmallID data-set

1. Fully convolutional networks vs. BLSTM-RNN
   - △ Greatly reduce training time (on GPU)
   - △ Competitive accuracy and CER
   - △ Divide inference time by three on CPU
2. Dilation depth improves the result but with additional storage and time cost
3. Flat designs fails to reach good performance which confirms the need to use different levels of dilation
4. Skip or residual connections are proved to be an improvement above standard Pyra-DCNN

Experimentation & Results

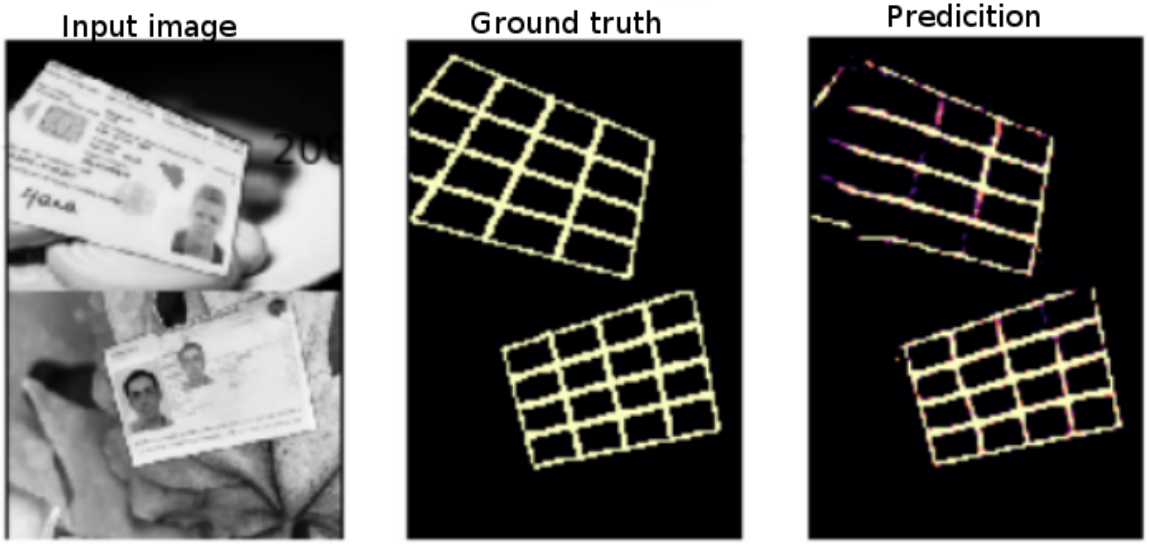
ARIADNEXT
WE DO IDENTITY.

Detailed results on BigID Data-set

model	**Accuracy** (%)	**CER**	**inference time** (in ms on CPU)	**Number of weights** (x 1000)	**training time** on GPU
PyraD-DCNN	**94.65**	**0.81**	13	454	13h20
LSTM-128	94.20	0.84	44	360	50h
Deep PyraD-DCNN	**94.99**	**0.74**	14	627	14h50m
Shallow PyraD-DCNN	93.97	0.91	12	109	12h10m
PyraR-DCNN	92.83	1.11	13	300	16h20m
1-Flat-DCNN	93.68	0.97	13	454	20h30m
Block PyraD-DCNN	93.82	0.95	12	421	13h30m
UpUp-DCNN	94.51	0.82	13	454	13h40m
DownDown-DCNN	94.37	0.85	13	454	13h20m

Discussion

- Up-Down design of the PyraD-DCNN is closely related to auto-encoder architectures (like U-Net)
- This network is also compatible with other tasks such as object detection or image labelling (ex. Document localization)



Conclusion and perspectives

ARIADNEXT
WE DO IDENTITY.

- Dilated convolutions based auto-encoder networks could replace traditional BLSTM-RNN layers
- Lightweight and more computational efficient on CPU which opens opportunities for mobile applications
- Wide range of applications such as for end-to-end scene text detection and recognition, sound or video processing

Thank you for your attention!