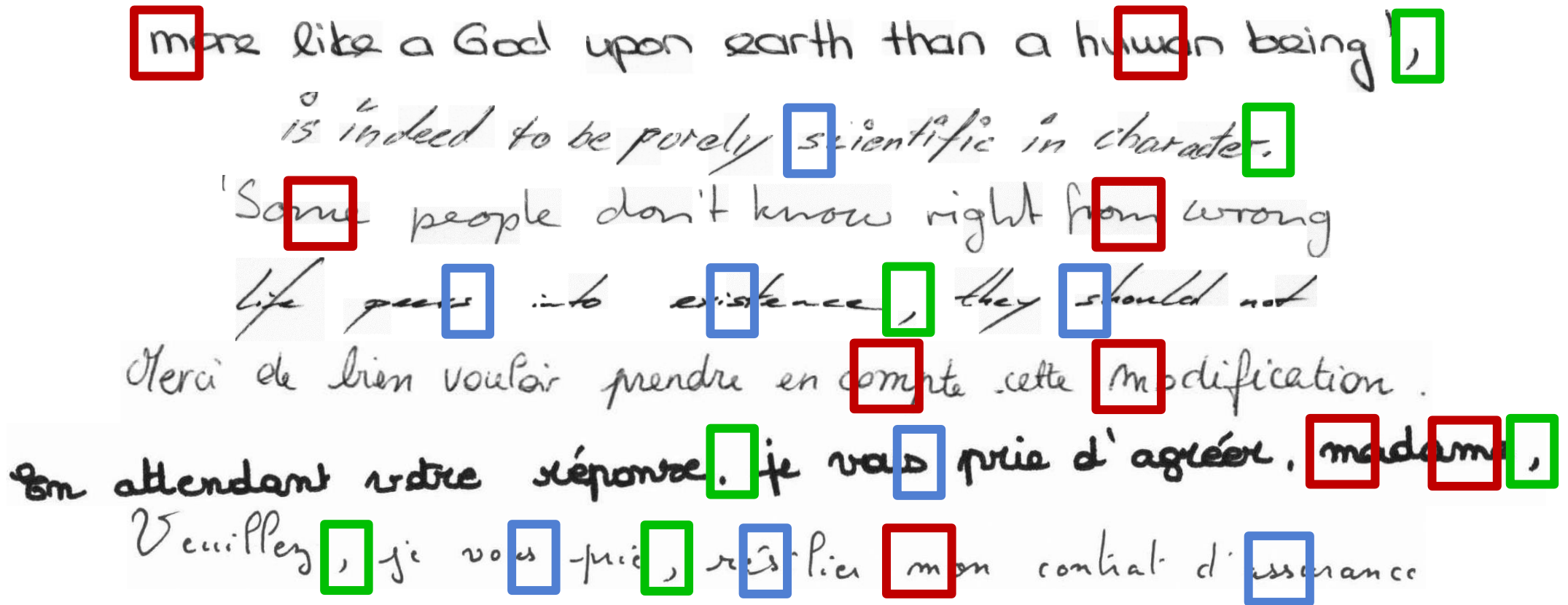


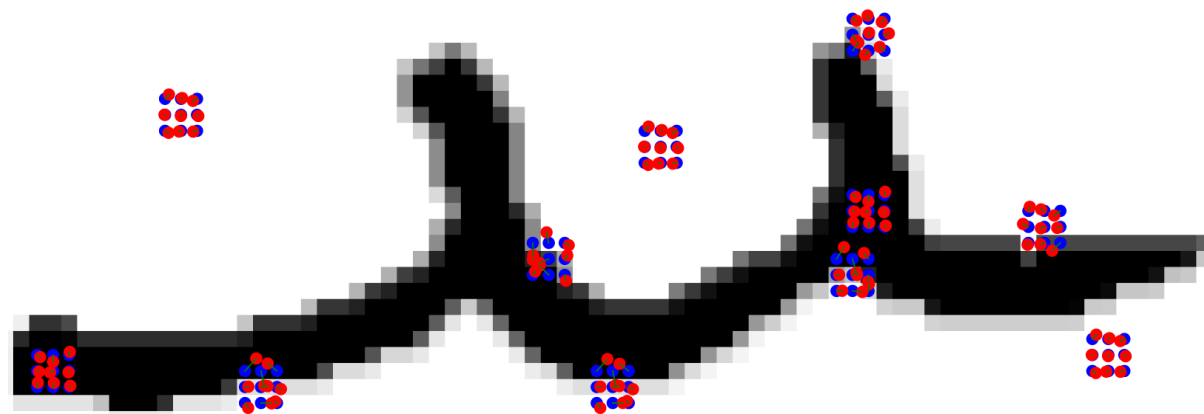
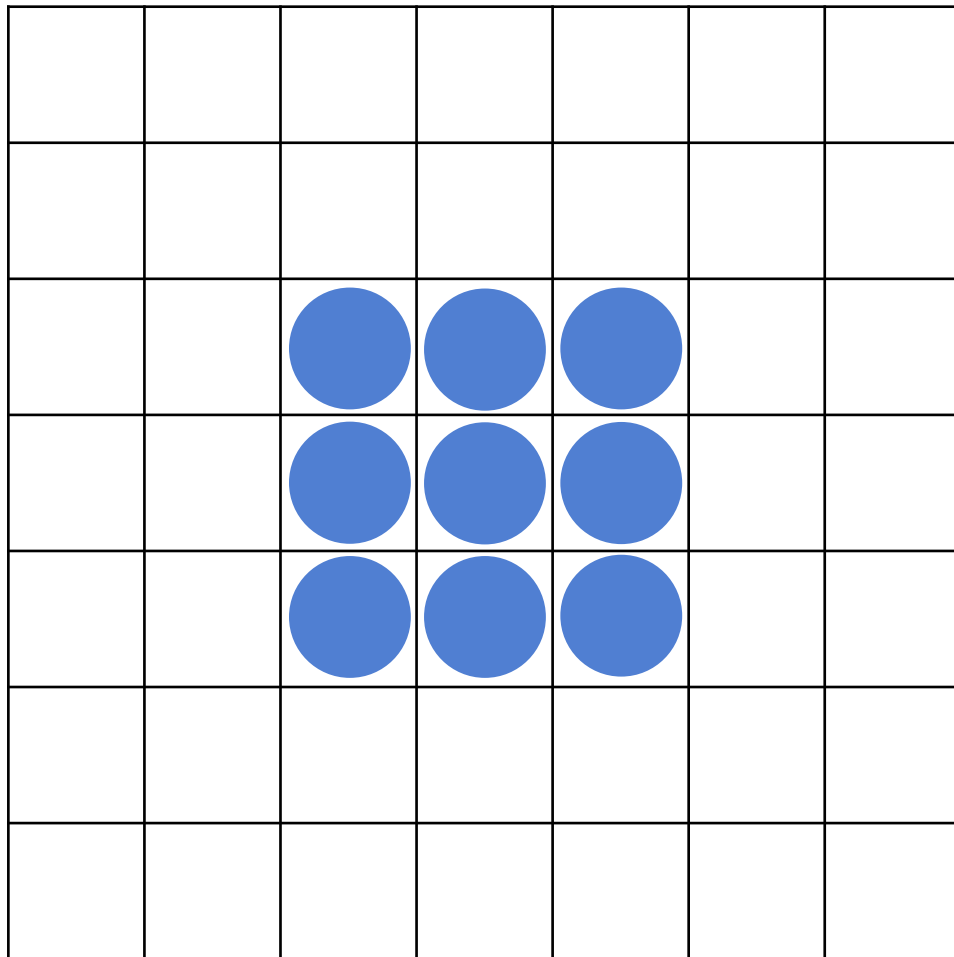


Handwritten text is a sparse structure, and characters and words vary in shape, scale, and orientation

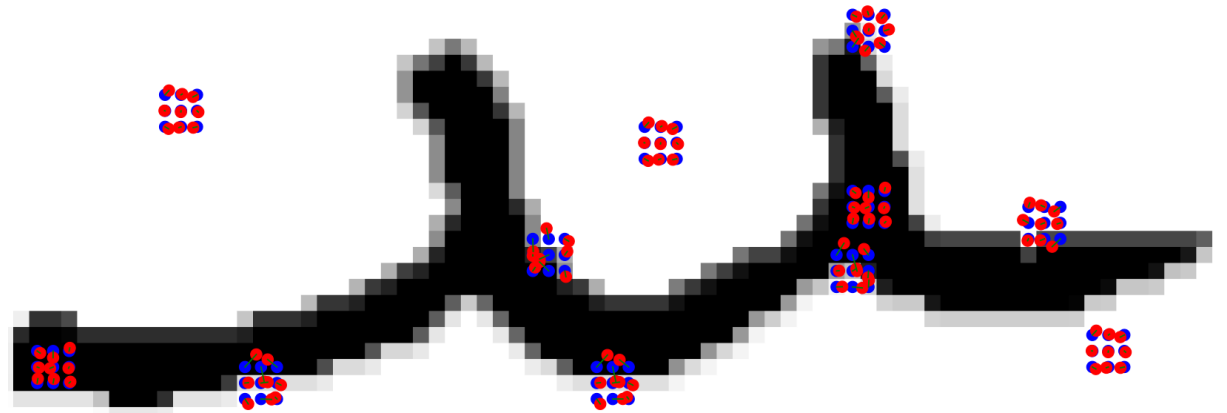
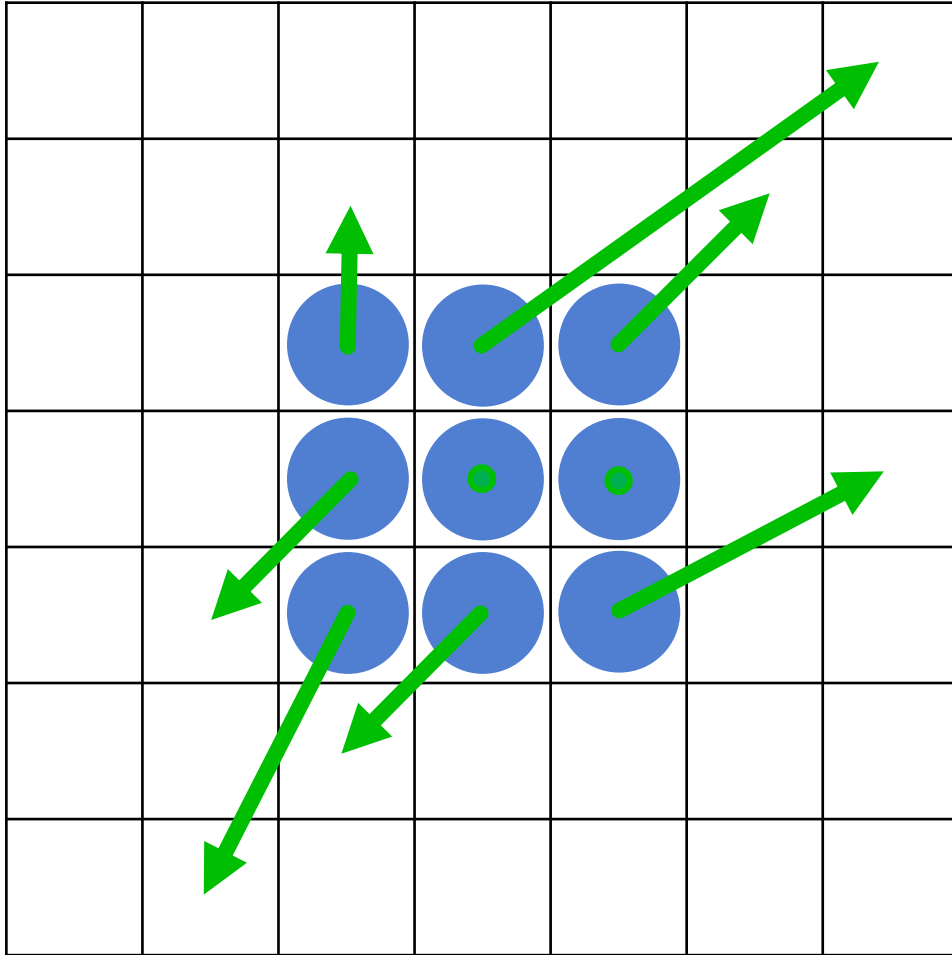


The image displays four lines of handwritten text in various scripts, with individual characters and words highlighted by colored bounding boxes (red, green, and blue) to illustrate their irregular shapes, scales, and orientations. The first line is in English: "more like a God upon earth than a human being,". The second line is also in English: "is indeed to be purely scientific in character.". The third line is in English: "Some people don't know right from wrong life goes into existence, they should not". The fourth line is in French: "Merci de bien vouloir prendre en compte cette modification. En attendant votre réponse, je vous prie d'agréer, madame, Veuillez, je vous prie, recevoir mon cordial d'assurance".

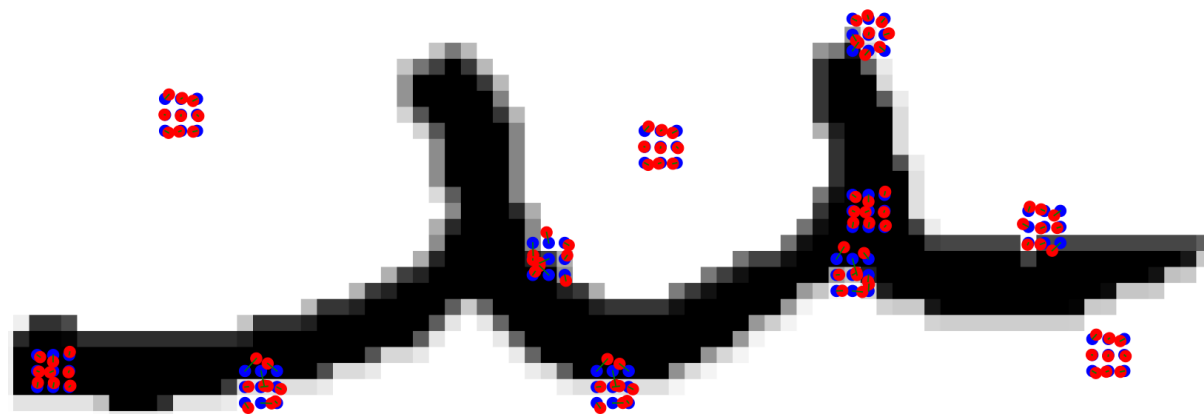
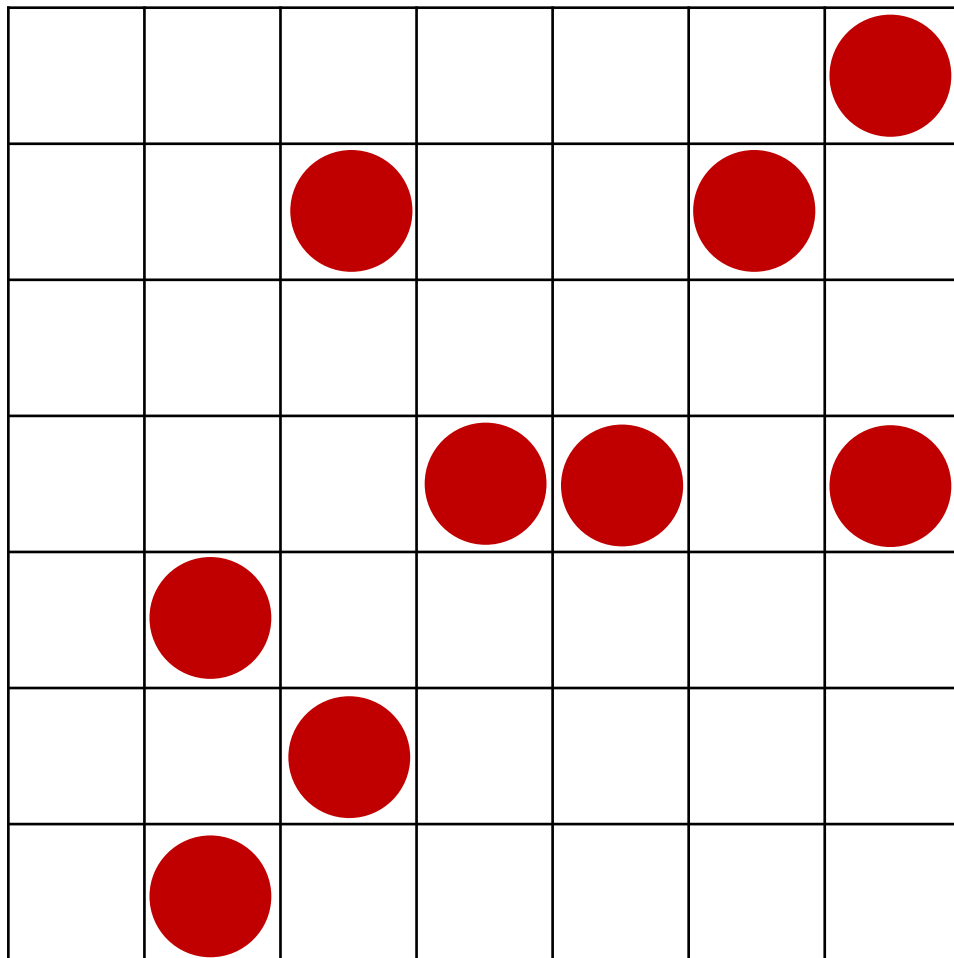
→ We take into account these characteristics by using DefConvs in place of standard convolutions for HTR



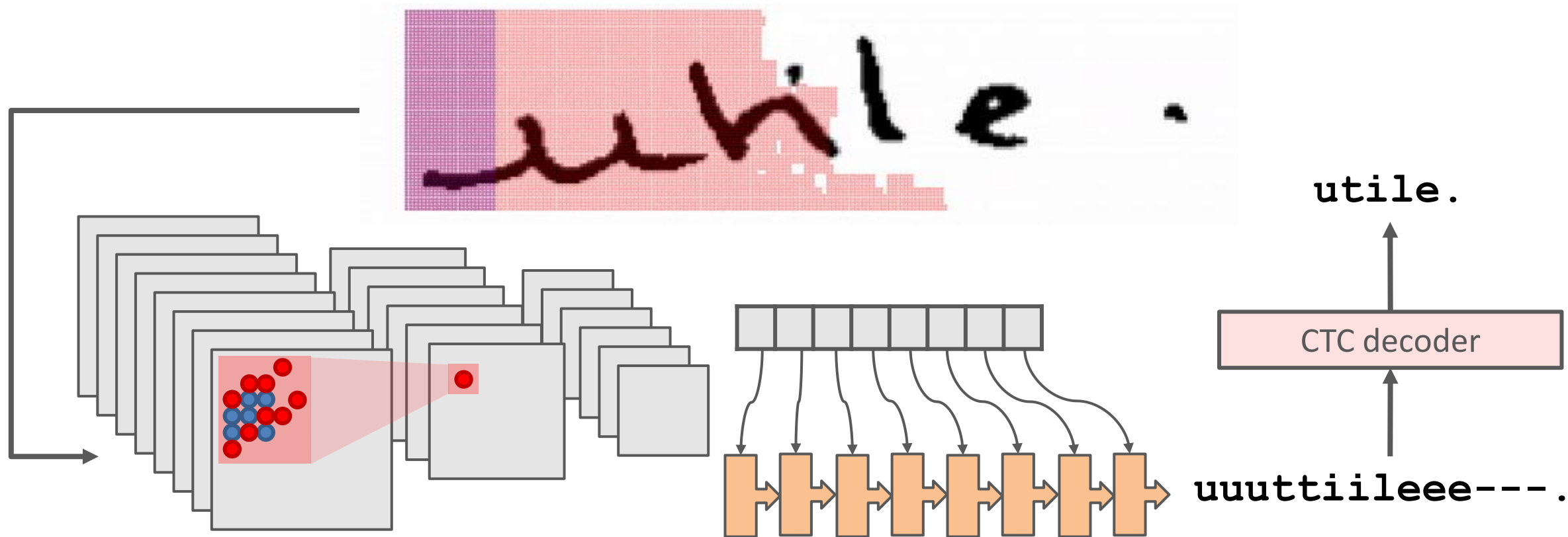
$$y(p) = \sum_{d \in \mathcal{N}} (d) \cdot x(p + d)$$



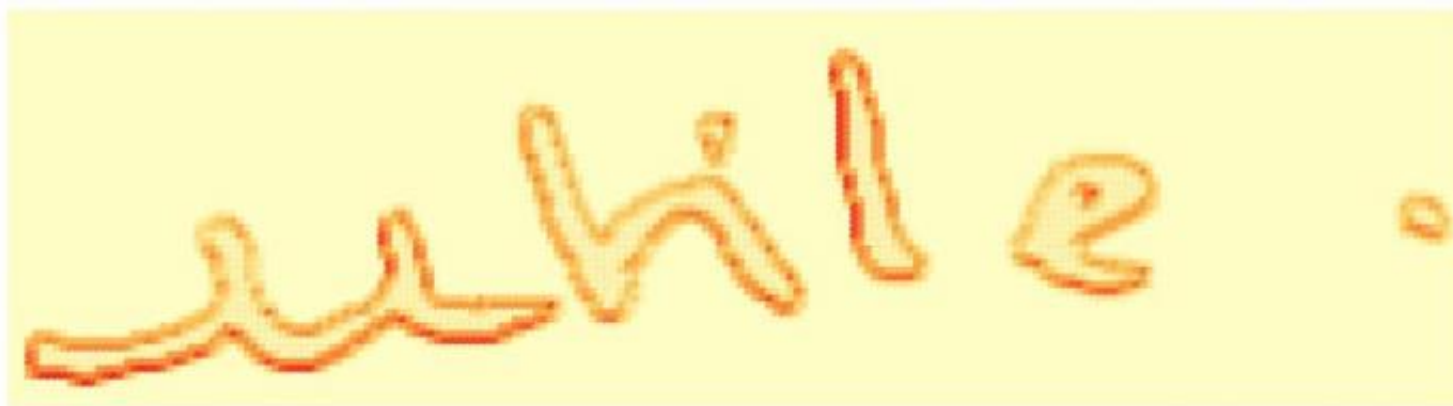
$$y(p) = \sum_{d \in \mathcal{N}} (d) \cdot x(p + d + \Delta d)$$



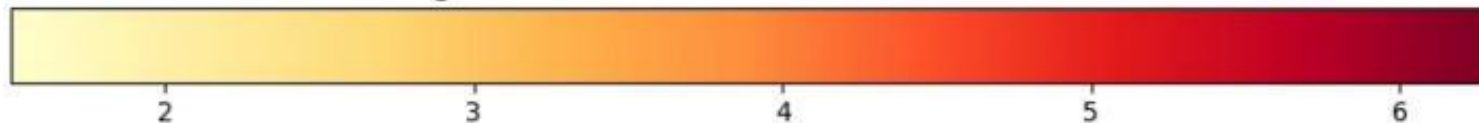
$$y(p) = \sum_{d \in \mathcal{N}} (d) \cdot x(p + d + \Delta d)$$



 The best friend of the family, love the



Cumulative offsets magnitude

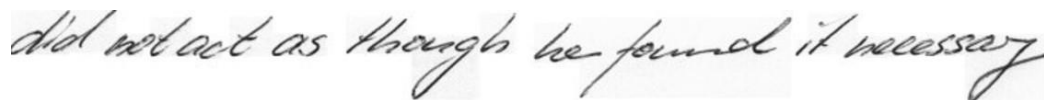


	IAM dataset		RIMES dataset	
	CER	WER	CER	WER
Full-DefConv	4,6	19,3	4,6	14,8
Shi <i>et al.</i> [2]	5,7	23,2	5,3	17,5
Wigington <i>et al.</i> [3]	6,4	23,2	2,1	9,3
Voigtlaender <i>et al.</i> [4] – LM	8,3	27,5	4,0	17,7
Puigcerver [5]	6,2	20,2	2,6	10,7
Bluche [6]	7,9	24,6	2,9	12,6
Pham <i>et al.</i> [7]	10,8	32,1	6,8	28,5

- [2] B. Shi, X. Bai, and C. Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Trans. PAMI*, 39:2298–2304, 2016.
- [3] C. Wigington, C. Tensmeyer, B. Davis, W. Barrett, B. Price, and S. Co-hen. Start, follow, read: End-to-end full-page handwriting recognition. In *ECCV*, 2018.
- [4] P. Voigtlaender, P. Doetsch, and H. Ney. Handwriting recognition with large multidimensional long short-term memory recurrent neural networks. In *ICFHR*, 2016.
- [5] J. Puigcerver. Are multidimensional recurrent layers really necessary for handwritten text recognition? In *ICDAR*, 2017.
- [6] T. Bluche. Joint line segmentation and transcription for end-to-end hand-written paragraph recognition. In *NeurIPS*, 2016.
- [7] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour. Dropout improves recurrent neural networks for handwriting recognition. In *ICFHR*, 2014.



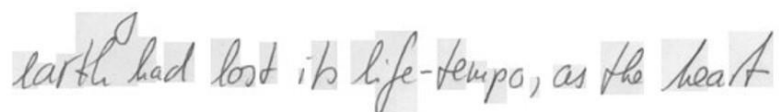
## On the IAM dataset:



**Groud Truth:** did not act as though he found it necessary

**Full-DefConv:** did not act as though he found it necessary

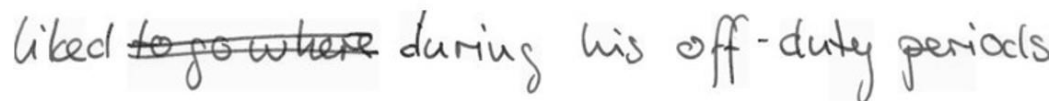
**Shi et al. [2]:** dd n act as thaugh kefanod it necarseay



**Groud Truth:** earth had lost its life-tempo, as the heart

**Full-DefConv:** earth had lost its lefe-tempo, as the heart

**Shi et al. in [2]:** earthled bost its eferteupo, as the beat



**Groud Truth:** liked during his off-duty periods

**Full-DefConv:** liked # during his off-duty periods

**Shi et al. [2]:** liked tegotere during his off-duty periots

## On the RIMES dataset:



**Groud Truth:** Md Dubois je souhaiterais être couvert au titre de la responsabilité

**Full-DefConv:** Ma Duois je souhaiterais être coupet au. Titre de l ressonabilité

**Shi et al. in [2]:** ma Buboiss. Je souhatersir être lea mert u titre de b ressonssbilité



**Groud Truth:** JE ME PERMETS DE VOUS ECRIRE POUR AVOIR

**Full-DefConv:** JE ME MERMETS DE VOUS ECPIR POUP QNOIR

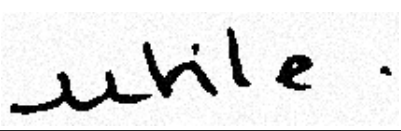
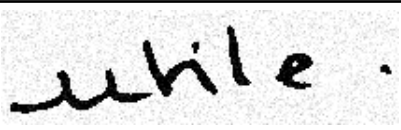
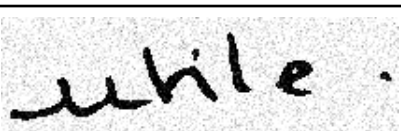
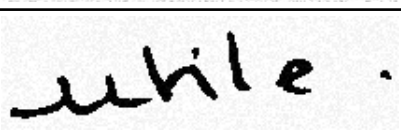
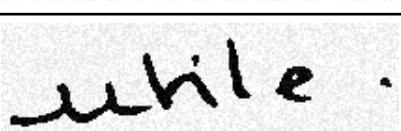
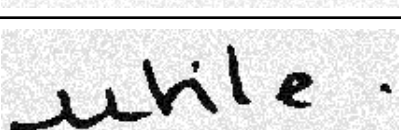
**Shi et al. in [2]:** JEe PERMET DE voUS FcAlrRe Pour avoin



**Groud Truth:** merci de votre collaboration

**Full-DefConv:** erci de votre collaloration

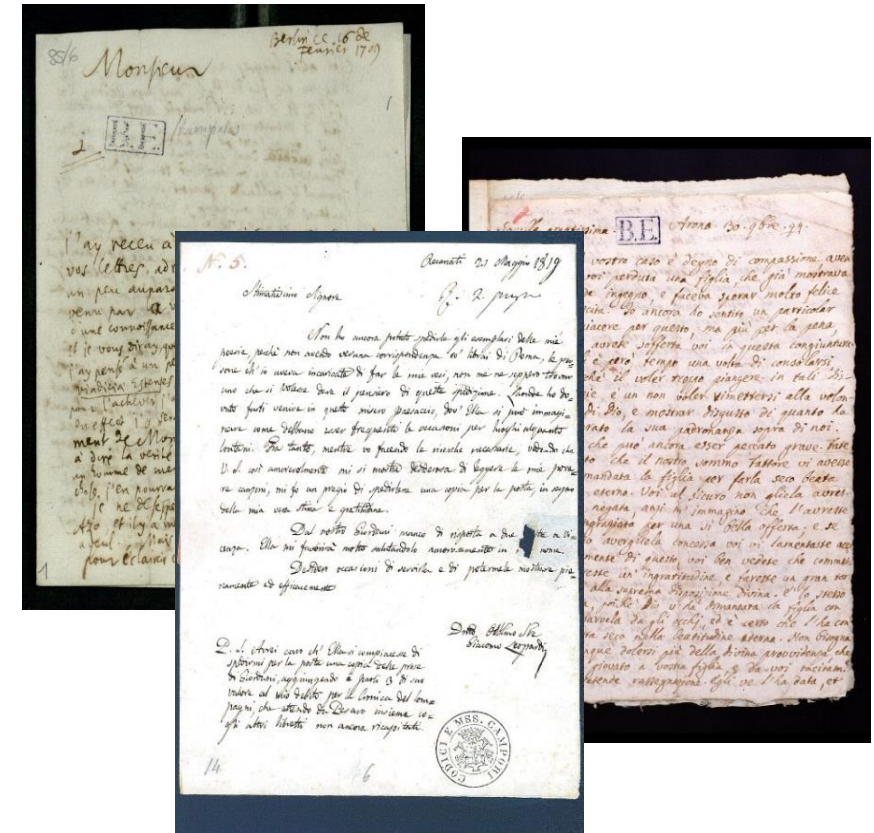
**Shi et al. in [2]:** A'acexio de votre collaloration

		IAM dataset				RIMES dataset			
		Full-DefConv		Shi <i>et al.</i> [2]		Full-DefConv		Shi <i>et al.</i> [2]	
		CER	WER	CER	WER	CER	WER	CER	WER
$\mathcal{G}(0,10)$		4,7	19,5	5,8	23,7	4,6	14,8	5,3	17,3
$\mathcal{G}(0,20)$		5,5	22,2	6,9	26,5	4,7	15,4	5,4	18,2
$\mathcal{G}(0,30)$		8,3	49,0	24,4	62,8	5,1	17,0	6,0	20,2
$\mathcal{P}(0,10)$		4,8	19,8	5,9	24,0	4,6	14,8	5,3	17,4
$\mathcal{P}(0,20)$		5,5	22,0	6,7	26,0	4,6	15,1	5,4	17,7
$\mathcal{P}(0,30)$		10,6	33,3	13,6	41,2	4,7	15,1	5,5	18,2

We showed that DefConvs are more suitable than standard convolutions for the task HTR task since they

- Can adapt to highly distorted handwritten strokes
- Are robust to background noise

→ These features make them promising for being applied for HTR of historical manuscripts (both from benchmark datasets and new ones that we are currently collecting)



# Thank you!

## Watch Your Strokes: Improving Handwritten Text Recognition with Deformable Convolutions

Iulian Cojocaru, Silvia Cascianelli, Lorenzo Baraldi, Massimiliano Corsini, Rita Cucchiara

For questions and discussions, please “drop by” our poster or reach out to Silvia Cascianelli: [silvia.cascianelli@unimore.it](mailto:silvia.cascianelli@unimore.it)