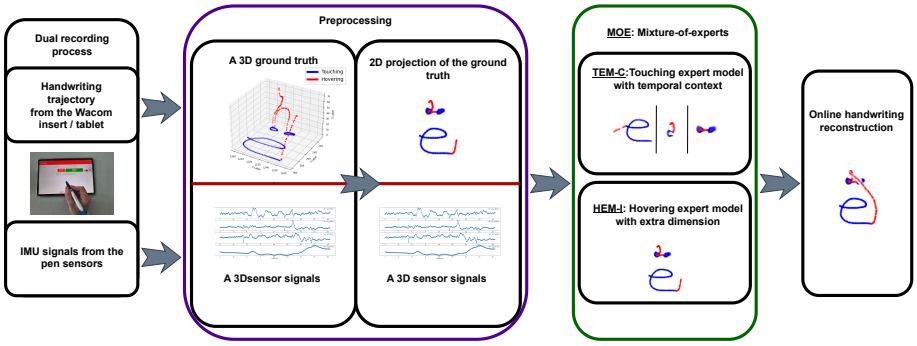
Mixture-of-experts for handwriting trajectory reconstruction from IMU sensors

Florent Imbert, Eric Anquetil, Yann Soullard, Romain Tavenard

Graphical Abstract

Mixture-of-experts for handwriting trajectory reconstruction from IMU sensors

Florent Imbert, Eric Anquetil, Yann Soullard, Romain Tavenard



Highlights

Mixture-of-experts for handwriting trajectory reconstruction from IMU sensors

Florent Imbert, Eric Anquetil, Yann Soullard, Romain Tavenard

- a mixture-of-experts approach made of two expert neural networks, one for the touching trajectory, the other for the hovering parts of the trajectory;
- integration of a temporal context reflecting physics and dynamics to improve the expert model dedicated to the touching part, and is also intended to facilitate joining with the other expert model;
- taking into account the extra dimension present in the hovering parts linked to the height of the pen to improve the expert model dedicated to the hovering part;
- a mixture-of-experts, with each expert optimized both on it's dedicated task and to facilitate collaboration with the other expert, to obtain a mixture of experts forming a coherent and robust system;
- a new public database that will serve as a benchmark to help advance research.

Mixture-of-experts for handwriting trajectory reconstruction from IMU sensors

Florent Imbert[a], Eric Anquetil[a], Yann Soullard[b], Romain Tavenard[b]

[a]*IRISA, Universite de Rennes, INSA Rennes, Rennes, France*
[b]*IRISA, Universite Rennes 2, Rennes, France*

Abstract

The use of digital pens for online handwriting trajectory reconstruction is a prevalent method for human-computer interaction. In this study, we focus on a digital pen equipped with sensors where we aim at reconstructing the online handwriting trajectory. This pen enables writing on any surface and preserving the digital trace of handwriting. This type of pen could be used as an aid to learning to write in classroom. In this paper, we propose a new approach learning to finely reconstruct the touching trajectories while precisely analyzing the hovering part in order to position the next touching trace correctly. This relies on a Mixture-Of-Experts (MOE) approach. The first expert is dedicated for the pencil touch, and is named touching expert model. The second one is dedicated for the hovering pen trajectory, and is named hovering expert model. We improve on the learning of each of these experts based on additional context or specific examples. In addition we introduce a novel public benchmark dataset, to enable future research and comparisons in the field of handwriting reconstruction. The results demonstrates a significant enhancement compared to its primary competitors

Keywords: Online Handwriting, Trajectory Reconstruction, Digital Pen, Inertial Measurement Units, Deep Neural Network

1. Introduction

Digital devices play a crucial role in enhancing the learning experience for both students and teachers by facilitating active learning techniques and offering immediate

Email address: florent.imbert@irisa.fr (Florent Imbert)

feedback [1]. The literature on e-learning [2, 3, 4] highlights the accuracy and reliability of computer-based analysis for generating relevant feedback for correction or guidance. Based on this, pen-based tablet applications have been developed to provide personalized feedback [5, 6].

Despite the increasing reliance on digital platforms, there remains a need for children to learn writing on paper, as it remains the most widely used surface. To address this, digital pens, such as the Digipen stylus developed by STABILO, have been equipped with kinematic sensors (Inertial Measurement Units) to track pen movements (Fig. 1). Such a pen allows to capture handwriting gestures on any surface, including paper, and it allows to visualize and analyze the handwriting reconstruction on a digital medium (e.g. tablet, computer).

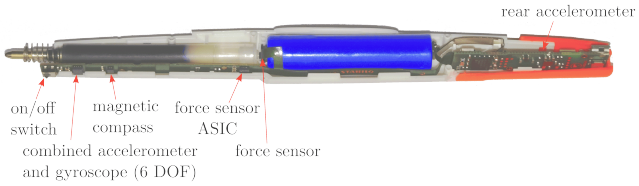


Figure 1: © STABILO International Digipen's sensor location

This work focuses on the challenging task of reconstructing the digital handwriting trajectory from the Digipen stylus. While Inertial Measurement Units (IMUs) are commonly used in tracking systems due to their cost-effectiveness, their inaccurate signals due to high noise levels are problematic. Unlike handwriting recognition tasks, where a label exists for the global shape of the handwriting, trajectory reconstruction involves a precise reconstruction at each time frame by analyzing local displacement.

This research hypothesizes that the precise reconstruction of handwriting trajectories from IMU data can significantly improve feedback mechanisms in educational applications. This enhances both learning outcomes and user experience, while preserving the natural experience of writing on paper. By enabling real-time handwriting instruction and correction, this approach offers the dual advantage of writing on physical paper and benefiting from digital analysis and feedback, helping to make e-learning tools more effective. Thus, the designed system must produce the most accurate tra-

jectory reconstruction possible for e-learning applications, where precise feedback to learners is essential.

Previous studies have utilized IMU sensors for various applications, including recognizing predefined movements [7, 8, 9], reconstructing pedestrian trajectories [10], and estimating upper limbs kinematics during industrial gestures of pick and place [11]. Several previous works investigated handwriting reconstruction using the Digipen stylus. [12] propose a Convolutional Neural Network for this task. This extends the works of [13] to multiple writers. [14] is the third attempt to reconstruct handwriting trajectories from the Digipen using deep neural networks. It has produced promising results in terms of handwriting reconstruction. Their approach is dedicated to the reconstruction of touching parts, which are the only parts for which a reliable ground truth is available during training.

IMU sensors, while useful for tracking motion, are prone to accumulating errors over time, a phenomenon known as sensor drift. In the field of handwriting trajectory reconstruction, prior methods [15, 16] have faced significant limitations due to sensor drift. This drift causes imprecision in long-duration or continuous handwriting tasks, making it difficult to reconstruct accurate trajectories. In particular, previous approaches [12, 17] that treated handwriting as a single continuous motion without segmenting the writing into distinct parts often amplified the effects of this drift, leading to inaccurate global reconstructions.

To address this issue, we propose a novel handwriting trajectory reconstruction approach consisting of two expert networks to deal with the different nature of IMU signals. Our proposal is based on the key idea of separating the task into two parts: reconstructing the handwriting trajectory when the stylus touches the surface and when the pen is up, which is called the hovering trajectory. In fact, touching trajectories are 2-dimensional trajectories referring to the writing itself and where ground truth can be obtained using double Digipen-Wacom acquisitions (Fig. 5). Hovering trajectories are 3-dimensional trajectories corresponding to a target trajectory towards the next part of the writing and where ground truth signals can only be recovered under a 7mm height. This is an additional difficulty to take into account when dealing with hovering trajectories and that, to our knowledge, has not been properly considered in the literature.

Thus, in this work, the goal is twofold. First, each type of movement (2D touching and 3D hovering parts) has its own dynamic properties, and our approach assigns specialized expert models to handle these segments independently. Second, by separating the two tasks, the proposed approach aims to reduce the cumulative error introduced by sensor drift. This motivates our choice of designing a mixture-of-experts approach with one expert neural network for each of these specific tasks.

Our main contributions are presented below:

- a mixture-of-experts approach made of two expert neural networks, one for the touching trajectory, the other for the hovering parts of the trajectory;
- the integration of temporal context into the expert model dedicated to the touching part, reflecting physics and dynamics and facilitating the consistency with the other expert model;
- Enhancing the capabilities of the expert model dedicated to the hovering part by taking into account the additional dimension due to the height of the pen during training;
- a mixture-of-experts, where each expert is specialized on its dedicated task and where the collaboration with the other expert is facilitated, to obtain a mixture of experts forming a coherent and robust system;
- a new public database that will serve as a benchmark to help advance research.

The training and testing phases are described in details, to enable reproducibility of the experiments relying on an open database.

The rest of the paper is organized as follows, related works are presented in the following section 2. A short review of the metrics used for handwriting evaluation is presented in section 3. Section 4 introduces our innovative approach to expert neural networks dedicated to each part of writing (touching and hovering parts). We then present improvements for each expert. The experimental results are presented and discussed both quantitatively and qualitatively in section 5. An ablation study shows the impact of each contribution. Finally, we conclude and discuss perspectives.

2. Related works

The field of digital devices for note-taking, drawing, and handwriting learning is rapidly expanding. However, most systems use display for digital handwriting acquisition, with only a few incorporating styluses equipped with motion tracking systems to reconstruct handwriting trajectories. In this section, we focus related works on handwriting trajectory reconstruction and the use of Inertial Measurement Unit (IMU) sensors for handwriting recognition.

2.1. Handwriting Trajectory Reconstruction

There are various systems dedicated to the acquisition of digital handwriting. Most of them use pen-based tablets with display, such as the Samsung S Pen, Apple Pencil, Microsoft Surface Pen, and Wacom offering precise tracking but dedicated to specific devices. For these systems, the pen is powered by the electromagnetic field generated by the Electro-Magnetic Resonance (EMR) sensor located under the diaplay. Another approach enables writing on paper using pens equipped with tools to capture handwriting gestures. This can involve embedding cameras in styluses, as seen in Anoto pen, or employing IMU sensors as in the Digipen [18]. IMU-equipped styluses are versatile, allowing writing on tablets, paper, or boards. However, IMU signals only provide relative pen displacements, introducing potential noise.

Three categories of handwriting trajectory reconstruction emerge: i) reconstruction from offline handwriting images as [19, 20]; ii) reconstruction from pen-tip optical tracking systems as [13]; iii) reconstruction from IMU signals. Few works focus on IMU-based trajectory reconstruction. Some approaches [21, 13, 12], use deep learning for online handwriting trajectory reconstruction as a step toward handwriting recognition. Earlier methods, like those in [16], relied on traditional approaches like Hidden Markov models.

Recent advancements in handwriting trajectory reconstruction and IMU signal enhancement have highlighted several innovative approaches. [21] developed a handwriting assistant system that achieves millimeter-level accuracy by utilizing attachable inertial sensors, demonstrating the high precision achievable with IMU-based methods. Complementing this, [16] proposed a noise reduction technique specifically for

IMU signals, which could significantly enhance the robustness of handwriting trajectory reconstruction models by addressing sensor noise challenges. In a related study, [10] introduced an unsupervised method for reconstructing pedestrian trajectories from IMU data, whose methodologies could be adapted to improve handling noisy signals during hovering strokes in handwriting applications. Furthermore, [22] presented a wavelet encoding network that enhances inertial signals, providing valuable insights for improving signal fidelity in handwriting datasets. [23] also contributed with a GAN-based approach for inertial signal enhancement, which could inspire enhancements in reconstruction accuracy, especially for complex handwriting trajectories. [24] proposed a method for extracting and denoising high-resolution vehicle trajectories from aerial videos, significantly improving the precision of trajectory data in intelligent transportation systems. Lastly, [25] focused on ship trajectory reconstruction from AIS sensor data, successfully integrating deep learning techniques with wavelet analysis for optimal signal denoising and demonstrating effective applications in trajectory reconstruction tasks. Together, these studies represent significant strides in the field, offering methodologies and insights that could greatly benefit handwriting trajectory reconstruction efforts. [26] explored magnetic signals, while [12] addressed the Stabilo Digipen, using linear interpolation to align pen and tablet signals as preprocessing. An alignment method is necessary due to the requirement of matching sequence sizes for point-to-point loss calculation. They use a Convolutional Neural Network for the online handwritting trajectory reconstruction. Recently [14] improved on [12] by proposing a complete pipeline to achieve online handwritting trajectory reconstruction cf. Fig. 2. In particular, Dynamic Time Warping (DTW) is used to align the ground truth with the input signal, which has the advantage of preserving the dynamics of handwriting, unlike linear interpolation as used in [12].

They also propose the use of a Temporal Convolutional Networks (TCN) as a backbone network (Fig. 3), which has the advantage of having larger receptive field than a classical CNN as used by [12] .

[14] achieve very good results for the reconstruction of the touching parts of the handwriting, but the approach performs poorly for the reconstruction of the hovering parts. Based on the model from [14], [27] apply and compare different quantization

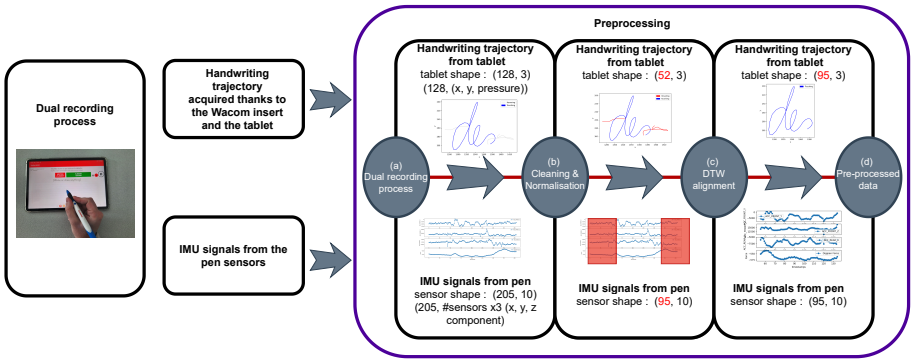


Figure 2: The processing pipeline proposed in [14]. (a) thanks to dual acquisition, they recover Digipen signals and the ground truth, (b) they remove start and end hovering, which are not data linked to handwriting, (c) they align ground truth and sensor signals using DTW, (d) preprocessed data used for training.

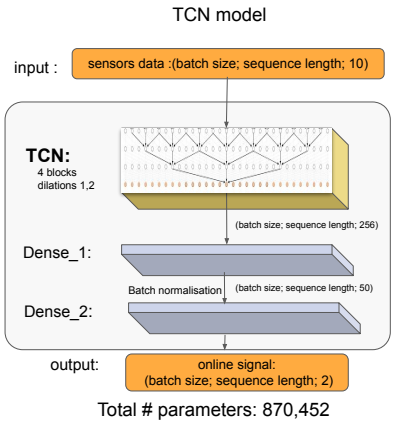


Figure 3: TCN model proposed by [14]

techniques in order to enable the on-device inference of handwriting trajectory regression from inertial data.

2.2. Using IMU Sensors for Handwriting Recognition

While few works focus on handwriting reconstruction from IMU sensors, online handwritting recognition from IMU sensor data has been explored using pen-tip trajectory signals and IMU sensor signals. It should be noted that, although the two

objectives may seem close, they are different in nature. The state-of-the-art shows that reconstructing from an IMU signal is a less complex task than reconstructing a precise trajectory. A benchmark study by [28] compared neural network architectures for character, symbol, word, and equation recognition from IMU signals. CNN/BLSTM architecture proposed in [29] showed the best results. Some works address both online handwriting trajectory reconstruction and character recognition. [12] trained two neural networks sequentially for these tasks, assessing reconstruction quality based on character recognition rates. [13] proposed multitask learning using a CNN-LSTM network, demonstrating improved trajectory reconstruction and character classification. In the context of the UbiComp 2021 Challenge, [30] utilized a CNN/BLSTM to recognize mathematical expressions written with the Stabilo's Digipen, focusing on label boundary quality. Other studies minimize the link bandwidth between the Digipen and the remote device (e.g. tablet) [31] or explore domain adaptation [32, 33] and explainability [34] in the context of online handwriting recognition from the Digipen. Recently, [35] use low cost IMU from smartphone to achieve characters recognition. Their specificity is to couple inertial dynamic signals with trajectory morphology information coming from the handwriting image, then use a CNN for classification.

3. Evaluation of Handwriting Reconstruction

Evaluation methods vary from one study to another, with some relying on qualitative assessments of reconstructed trajectories [36], while others use recognition performance as a proxy for evaluation [12, 37]. Metrics like Root Mean Squared Error (RMSE), Dynamic Time Warping (DTW) and Fréchet distance are also commonly employed to assess reconstruction accuracy [14, 19]. The diverse approaches underscore the complexity of evaluating handwriting trajectory reconstruction from digital stylus inputs.

In [14], the authors stated that the Fréchet distance is a good metric to evaluate the trajectory reconstruction. The Fréchet distance captures both local and global information accurately by taking into account temporal and spatial warping when aligning the predicted trajectory with the ground truth. It also correlates well with qualitative

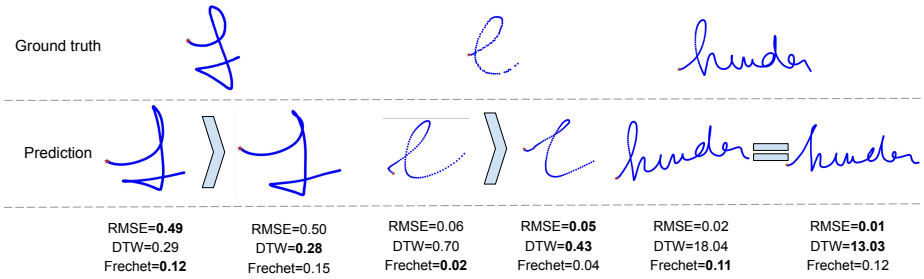


Figure 4: Comparison between Euclidean distance, DTW and Fréchet distance from predictions obtained using 2 different models. For the "f" and "e" characters, the predictions on the left seem better to us since the loops are better reproduced. For the word "hinder", the two predictions seem similar for us: the characters predicted on the left are better reproduced while the orientation is better on the right.

assessment in practice.

Our experiments confirm this observation, as shown in Figure 4. We have tested two experimental models and are observing the results for the different metrics. The Fréchet distance aligns most closely with visual perception, particularly in reconstructing loops, such as those in letters like "f" and "e". This is also the case for "hinder", where the two reconstructions are for us equivalent: in the first one, the loops and the characters are better made, but in the second one, the orientation is better. This results to close metric values, in contrast to the DTW values. Among various metrics, Fréchet distance uniquely captures the shape similarities. For this reason, we have chosen to only retain the Fréchet distance in our experiments.

As a reminder, for two multivariate time series $A \in \mathbb{R}^{T_A \times z}$, and $B \in \mathbb{R}^{T_B \times z}$ of equal feature dimensionality z and respective lengths T_A and T_B, the Fréchet distance is defined by the following equation:

$$F(A, B) = \min_{\delta} \max_{(i,j)\in\delta} d(a_i, b_j) \tag{1}$$

4. A new mixture-of-experts for better collaboration between 2D touching strokes and 3D hovering trajectories

We propose an original mixture-of-experts dedicated to handwriting reconstruction. We first justify our proposal and then present our two expert models, including the training strategies.

4.1. Motivation

Reconstructing handwriting trajectories from inertial measurement unit (IMU) signals is a complex task that requires advanced algorithms to handle the multidimensional nature of the data. Handwriting can be categorized into two phases: the act of writing (where the pencil makes contact with the surface) and the hovering phase (when the pencil transitions between two strokes). The touching part occurs in a two-dimensional plane as the act of writing is commonly on a flat surface while the hovering part is three-dimensional due to vertical variations of the pen. The hovering component presents particular challenges due to the additional dimension, leading to greater variability in trajectories depending on the individual. Moreover, this phase can exhibit increased randomness because gestures are not constrained by graphomotor movements, further complicating the reconstruction process.

Inspired by the work of [14], which demonstrated success in reconstructing touching strokes, we propose a new approach for dealing both with touching and hovering parts to reconstruct the entire trajectory of handwriting gestures. Our methodology is a mixture-of-experts that focuses on breaking the reconstruction process down into two tasks, the touching and hovering ones, resulting in a more nuanced understanding of each phase of handwriting.

The writing phase provides a two-dimensional ground truth based on the observable path of the pencil on the writing surface, making it amenable to direct measurement. In contrast, the hovering phase involves movement in three-dimensional space, during which the pencil must be positioned at the next stroke's starting point. This phase is challenging because it lacks a clear ground truth if the pencil is lifted too high. As depicted in (Fig. 5), the absence of precise 3D labels during training limits our ability to accurately model this phase.

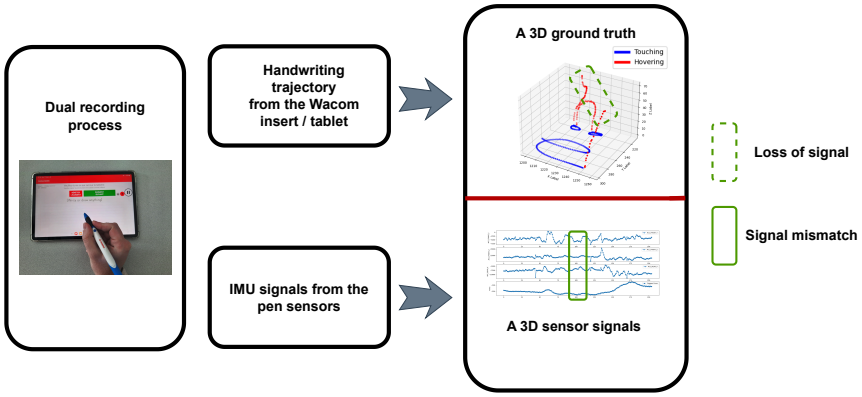


Figure 5: Dual recording process, with a Wacom insert, to enable the acquisition of ground truth. Composed of writing (blue) and hovering parts (red) that are not always tracked (green).

To address these challenges, our proposed MOE (Mixture-of-Experts) model effectively differentiates between the two phases of handwriting, allowing for targeted learning and improved trajectory reconstruction. By leveraging distinct approaches for the writing and hovering phases, our model aims to enhance accuracy while accommodating the variability inherent in human handwriting. To acquire ground-truth trajectory data corresponding to the IMU signals, we employ the Digipen stylus equipped with a Wacom insert as illustrated in Figure 5.

4.2. Our mixture-of-experts for handwriting reconstruction

In this section, we present the overall mixture-of-experts we propose and discuss in details the design of the two experts.

4.2.1. Presentation of our mixture-of-experts

We can formalize our problem as multitask learning, in that we have two linked tasks, the first being the prediction of the writing itself, the second the hovering movement between these different parts. These two tasks differ in their dynamics and nature (2-dimensional and 3-dimensional signals). Here, we propose to have 2 expert models, each specialising in its own task. Inspired by [14] where a TCN-based architecture is

trained on touching strokes only producing degraded reconstructions on hovering parts, we suggest to use this network architecture as a basis for our two expert models. In the following, we will refer to this TCN-based model as the backbone model.

We propose a new Mixture-Of-Experts (MOE) dedicated to handwriting reconstruction with two neural networks in parallel, one expert touching model and one hovering expert model. The mixture of models with the proposed training approach is illustrated in Figure 6.

The result is a global reconstruction of the handwriting. This approach functions as a mixture-of-experts model, with one expert focused on the handwriting during pen contact and the other handling the repositioning of the pen for the next stroke during hovering. The nature of the signals differs significantly between the constrained writing dictated by the graphomotor gesture and the repositioning trajectory during hovering. Unlike traditional MOE models, where switching between experts is predefined or data-driven, our model controls this transition dynamically through the pen's pressure sensor. This allows each expert to specialize more effectively, with the handwriting expert refining stroke precision and the hovering expert optimizing trajectory transitions.

We call this approach MOE-CI (Fig 6), which is the combination of the 2 following experts:

- the TEM-C for Touching Expert Model with hovering Context: corresponds to the backbone model trained with touching stoke with hovering portions preceding the touching strokes, it's an enhancement of the Touching Expert Model (TEM) which corresponds to the [14] method; it is dedicated to touching prediction;

- the HEM-I for Hovering Expert Model with Inclined data fine tuning, corresponds to the backbone model trained with complete sequence named Hovering Expert Model (HEM), followed by a fine tuning to data acquire on inclined plan; it is dedicated to the hovering prediction;

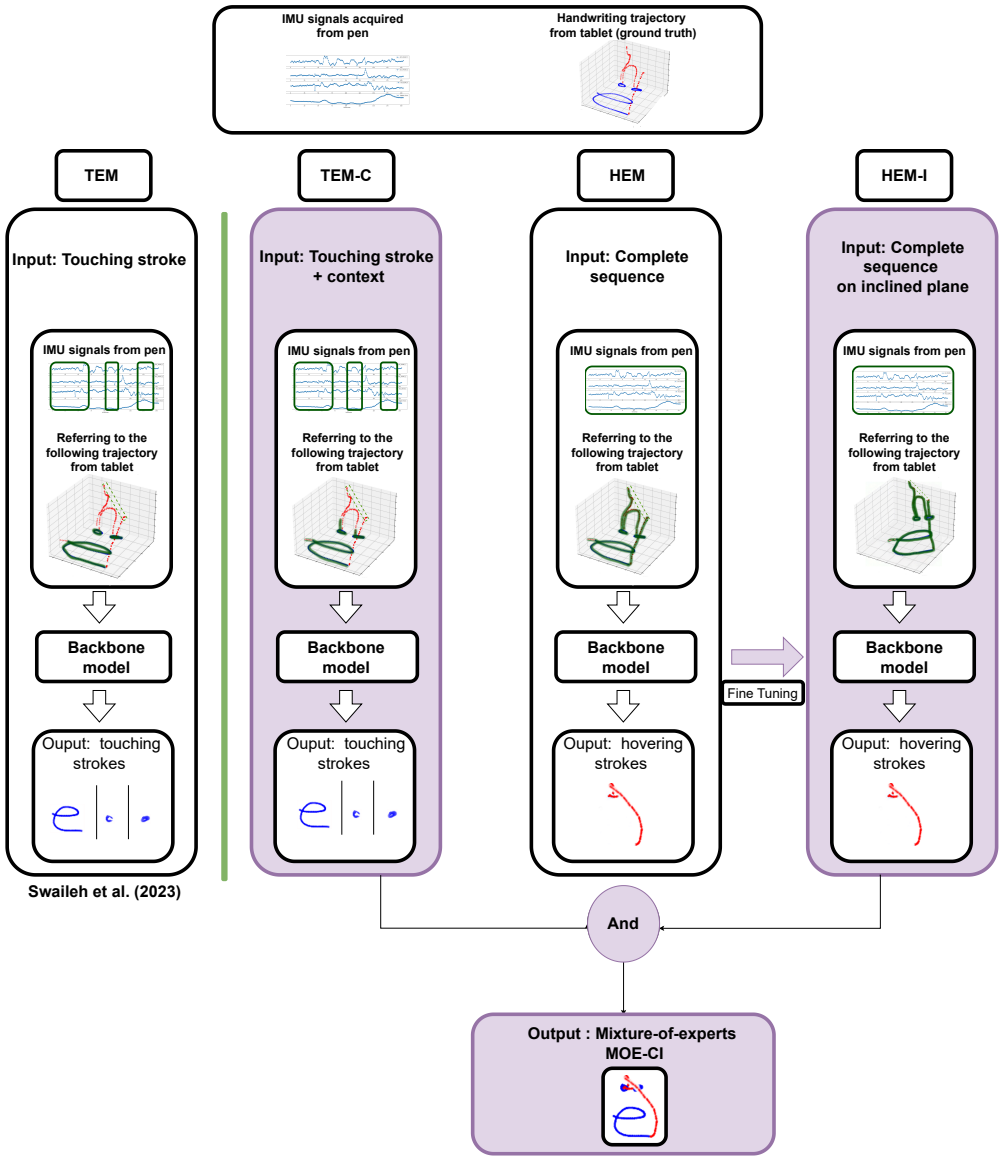


Figure 6: Our MOE-CI approach combines two models and their enhancements: TEM-C and HEM-I. TEM-C integrates hovering (in red) into the Touching Expert Model (TEM), providing better context and smoother transitions to touch. HEM-I fine-tunes the Hovering Expert Model (HEM) with 3D data for improved spatial understanding.

4.2.2. Network architecture

We remain convinced that a TCN-based network is the right network architecture to handle these IMU data. In addition, it has a number of advantages, including the fact that it has fewer vanishing gradient problems than LSTMs, is faster to train, and remains a lightweight architecture that requires considerably less data in training than a Transformer. This is particularly important in our use-case where the data is very specific and therefore relatively small in quantity. In addition, one of our final goal is to embed the model in the pen, which requires a lightweight architecture.

As a reminder and for reproductibility purposes, the backbone model (Fig. 3) is based on 4 blocks of a non-causal TCN followed by two dense layers, with a batch normalization layer applied between them. Each TCN block is composed of 2 convolutions with dilation 1 and 2 respectively, a kernel size of 3 and a dropout rate of 0.2. Thus, our two expert models, i.e. the touching expert and the hovering expert, are based on this network architecture. The network is trained conventionally, with the MSE as the cost function and early stopping at 25 epochs. 10 recordings are used for the test set, with the training data accounting for 90% of the remaining data and the rest for validation.

4.2.3. Models' inputs

We therefore turned our attention on how to train our models, in particular with regard to the type of data given as input to the network. Regarding the touching expert, the model is only trained on touching strokes, corresponding to 2D input signals. We believe that 3D signals from hovering parts may deteriorate the touching reconstruction due to the additional dimension that varies. Thus, the input signals are split into strokes according to the pen pressure and only the 2D signals from touching strokes are given in input of the touching expert model.

Addressing the specificity of hovering strokes, we suggest to use the backbone model on complete sequences, because we believe that giving as much context as possible can be beneficial for hovering prediction. The reason for training our network on entire sequences, rather than isolating hovering strokes, comes from the complex dynamics of pen-to-tablet interactions. By having the full sequence, the model gains

insights into the transition patterns between active stylus contact and hovering states.

Thus, TEM is trained on touching strokes, while HEM is trained on entire sequences. Additionally, we propose two variants: TEM-C and HEM-I, which differ in terms of input data. These variants are discussed in detail below.

4.3. TEM-C: Incorporating temporal context that reflects physics and dynamics to enhance the touching expert model

In the reconstruction of a trajectory based on Inertial Measurement Units, the temporal context plays a crucial role in accurately capturing the dynamic movement of an object. IMU are sensor systems that measure specific forces and angular rates to determine the acceleration and orientation of an object. Mathematically, the integration of these measurements over time helps to reconstruct object trajectory. The importance of temporal context in trajectory reconstruction process is the result from the necessity to accurately model continuous variations in acceleration and angular velocities, which reflect object dynamic behaviors including acceleration and directional shifts. By integrating temporal information, the object motion can be more accurately reconstructed, accounting for these dynamic changes.

This weakness has been confirmed in our experiments. The prediction was less accurate (Fig. 7) for the initial points of a stroke due to a lack of dynamics.

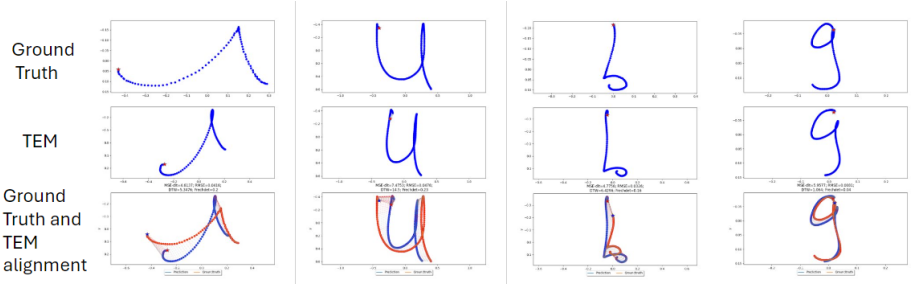


Figure 7: On the first line the ground truth. On the second the TEM [14] prediction. The observation is that the first points are less well reconstructed. On the last line the alignment between ground truth and the prediction.

To take this physical aspect into account, dynamic context is given to input during the network training phase. For that, hovering movements preceding the touching

strokes are added in the input sequences (in red on Fig. 8). The size of this hovering movements corresponds to half of the receptive field that can be captured by the model on the left border of the touching strokes, so that the model sees no padding to predict the positions associated with the first touching values. In addition, this will enable the network to see real signals instead of padding and thus to have a better generalization capability.

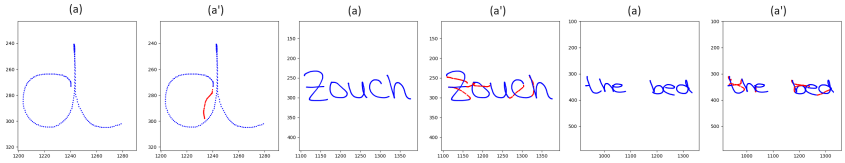


Figure 8: Visualizing the addition of hovering (in red) in (a') to ground truth in comparison to (a).

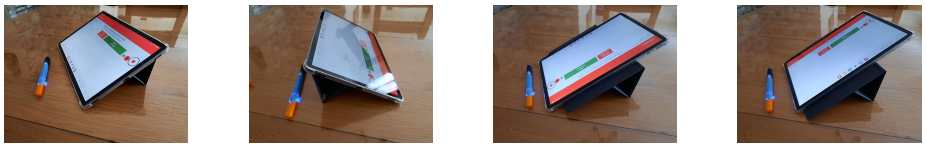
This effect on convolution padding can be translated mathematically as follows: let x be the input of a TCN convolutional layer, x is of dimension $T \times D^I$ where T is the sequence length, and D^I relates to the number of channels. Let f be a filter (convolutional kernel) of size $T^f \times D^I$. To preserve the input size in output, one considers a stride $s = 1$. Traditionally, when preserving the input size, the input sequence is padded on the borders by adding zeros on the left and on the right, which corresponds to $\lfloor \frac{R^f}{2} \rfloor$ zeros on the left and $\lfloor \frac{R^f}{2} \rfloor$ on the right added to the input sequence, where R^f corresponds to the receptive field of the TCN. Here, we introduce hovering part instead of padding to prevent padding effects on the touching prediction on the left border. We therefore have the following dilated convolutions:

$$y(i) = \sum_{k=-R_p}^{R_p} \sum_{d=0}^{D^I} f(k + R_p, d)\, x(i + r \times k, d) \tag{2}$$

where $R_p = \lfloor \frac{R^f}{2} \rfloor$ for a standard odd-sized kernel, r is the dilation rate and $y(i)$ is the output at position i. In a standard convolution, Equation 2 involves that $i + rk \geqslant 0$ ensuring valid indexing due to zero padding, whereas here, $i + rk$ is always valid due to the prior context from the pen's trajectory. This prior context is always available since there is always a downward trajectory of the pen before the first pressure point.

4.4. HEM-I: training on 3D labelled samples

Handling the hovering phase in trajectory reconstruction presents unique challenges due to distinct dynamics compared to writing segments. While the touching parts share a common 2D plane, hovering movements introduces a third dimension, representing the height of the hovering movements, adding complexity to sensor data as well as the variability of unconstrained hovering trajectories. To address this variability, we rely on a dedicated network, called Hovering Expert Model (HEM). This model shares the same architecture as our TEM. It is pretrained on data containing both touching strokes and hovering phases and fine-tuned using data acquired on inclined planes to benefit from variations in the 3 dimensions at training time. In this way, we expect that the fine-tuned network demonstrates enhanced adaptability to variations in hovering height, resulting in more robust predictions for hovering segments. Acquisition protocol to acquire inclined examples (Fig. 9) includes several positions to introduce variability in the inclination of the writing surface. Four setups are considered, positioning the tablet horizontally with a 30-degree upward or downward inclination, and vertically with similar 30-degree inclinations upwards or downwards.



(a) Horizontal tablet inclined upward

(b) Horizontal tablet inclined downward

(c) Vertical tablet inclined upward

(d) Vertical tablet inclined downward

Figure 9: Data acquisition protocol on inclined planes

5. Results

We experiment our approach on two datasets. One is public to be used for research work comparisons, the other is private with more data. First, we report a comparative analysis of our novel mixture-of-experts (MOE) against the established approaches of [14] and [12]. Then, we focus on the specific contribution of each expert model and

their impact on overall reconstruction through an ablation study. We perform a comparative exploration of possible mixtures, in order to show the benefit of the proposed improvements for each expert into the collaborative mixture-of-experts. Finally, experiments are done on both the private and public databases, highlighting the impact of data quantity and diversity.

5.1. Data sets description

In this work, we test our approach on two datasets, one public and one private. We can also note that these two datasets are extensions of those proposed by [14]. For this work we propose the KIHT-Public dataset, which contains 130 recordings. Each recording contains about thirty samples which are: characters, words, word groups, equations, and shapes (Fig. 10). The KIHT-Public dataset significantly surpasses the previous efforts by [14] in terms of recording quantity and diversity, featuring 130 recordings as opposed to only 30. This substantial increase allows for a greater number of unique writers, introducing more variability in handwriting styles. This is crucial for developing robust recognition models. Additionally, the inclusion of inclined recordings introduces further variability, particularly in pencil heights during writing, which can affect the dynamics of the strokes. This feature not only enriches the dataset but also makes it suitable for handwriting reconstruction tasks.

This dataset is publicly accessible[1] for research purposes. We have also evaluated our work on a private dataset (denoted KIHT-Private), which is an extension of the KIHT-Public one with 300 additional recordings. Details on datasets are available in Table 1. Note that the test recordings are the same for these two datasets to allow fair comparisons.

5.2. A new evaluation protocol

[14] highlighted that the Fréchet distance serves as an effective metric for evaluating trajectory reconstruction, which is why we have incorporated this metric into our evaluation. Given our focus on the shape rather than the size of the reconstruction,

[1] `https://www-shadoc.irisa.fr/irisa-kiht-s-and-kiht-public-datasets/`

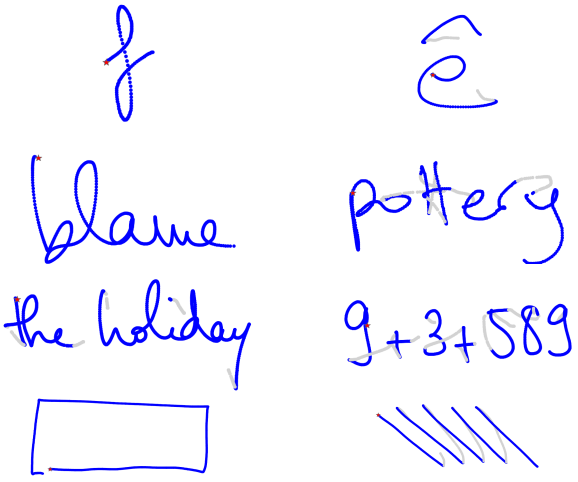


Figure 10: Some examples from the KIHT-Public dataset. Pen-up (hovering) strokes which are tracked by the tablet (i.e. under 7mm) are in gray and pen-down (touching) strokes are in blue.

Table 1: Important stats of the considered datasets

Datasets	Sets	# Writers	# Recordings	# Samples
KIHT-Public	Training	36	90	2761
	Inclined data	7	30	1368
KIHT-Private	Training	66	371	11811
	Inclined data	12	49	2234
Common test set		9	9	266

we propose two additional steps before calculating the Fréchet distance, as depicted in Figure (Fig. 11). The first step is to find the longest dimension of the ground truth bounding box (resp. the reconstruction), and set its size to 1. The aim is not to give too much weight in the evaluation to the size of reconstruction, but to the overall quality of reconstructions. The second step consists in centering the centroids of the prediction and ground truth bounding boxes. Another difference with [14] is that the Fréchet distance is not normalized by the sequence length in our evaluation protocol, so we can have a quantitative analysis of the impact of sequence length on reconstruction quality.

We also propose a dual-level evaluation. The first one is named label level, here we're evaluating all the touching strokes of a label. This process allows us to assess the

overall quality of the reconstruction, both writing parts and repositioning errors. The second one is named stroke level and it refers to individual touching strokes only. It quantifies the quality of the handwriting reconstruction.

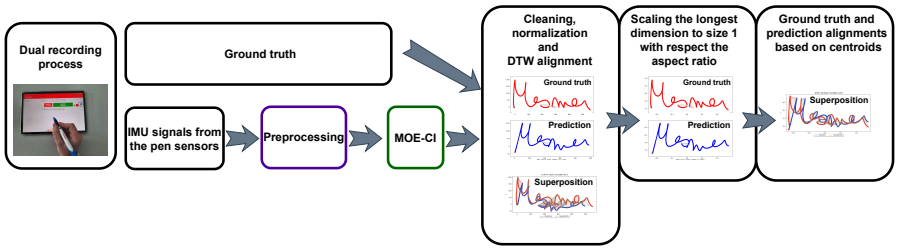


Figure 11: Our evaluation pipeline, composed of four steps, dual acquisition of Digipen signals and ground truth, alignment with DTW to find identical sequence sizes, scaling and centering on centroids.

5.3. Comparison of our mixture-of-experts (MOE-CI) with state-of-the-art methods

We compared our MOE-CI approach to the work of [14] and [12], the two references in the field. As a reminder, they proposed an approach based on a TCN model learned on touching strokes only for [14] and a CNN model with a linear interpolation between sensor signals and ground truth for [12]. We experiments these three approaches on the KIHT-Private dataset. These results are presented qualitatively (Fig. 12) and quantitatively (Table. 2). The Fréchet distances at label and strokes levels are significantly lower (and better) with our mixture-of-experts. We can see that on the touching parts, our approach reaches results close to those of [14]. One recalls that [14] has a model dedicated to touching strokes. A model specifically designed for hovering enhances the accuracy of predictions, as illustrated in Figure 12, where the repositioning between strokes is much more accurate than for the baseline method [14]. Additionally, incorporating contextual information into the touching model significantly improves the prediction of touching parts. This is evident at the stroke level of "faithful," where the loop in "ai" is better formed using the MOE-Ci method than with [14] approach. Overall, the MOE-CI allows each model to be tailored to the two different signal natures, leading to improved understanding and reconstruction.

Table 2: Fréchet distance computed for our mixture-of-experts and state-of-the-art methods train on the KIHT-Private dataset, and evaluate on the test set. The line "# wins" counts the number of labels where the model obtains the best scores against the others.

Evaluation	[12]	[14]	Our: MOE-CI
Label level	0.571	0.437	**0.312**
Stroke level	0.120	0.097	**0.091**
# Wins	24	54	**188**

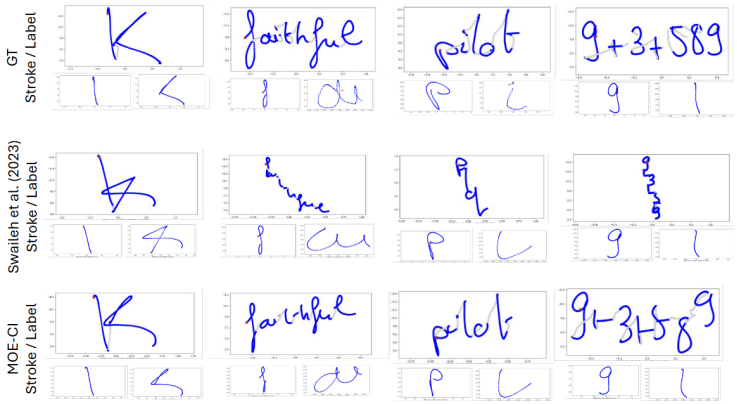


Figure 12: Comparison of our approach to that of [14], on the first line the ground truth at stroke (down) and label (up) level, on the second the reconstruction following the method of [14], on the last line our mixture-of-experts (MOE-CI).

5.4. Ablation study

In this section, we focus on the ablation study to evaluate the impact of different contributions on the overall performance of our mixture-of-experts. We first explore the impact of the temporal context integration on the touching expert (TEM vs TEM-C) (cf. 5.4.1), and then the impact of the extra dimension on the hovering expert (HEM vs HEM-I) (cf. 5.4.2). This ablation study is performed on the KIHT-Private dataset.

5.4.1. Temporal context integration on the touching expert (TEM-C)

As presented in section 4.3, we suggest to integrate temporal context in input to the Touching Expert Model. Results are shown at stroke level both quantitatively in Tab. 3 and qualitatively in Fig. 13. Adding temporal context significantly improves the model's ability to capture dynamic movement. By adding an earlier hovering portion

to the touching strokes inputs, the model can now account for past dynamics, providing a more comprehensive understanding of the trajectory. This can be seen in Figure 13, where the initial loop of the ”a” in ”faithful” is more accurately rendered, and the ”l” is more precisely shaped. Similarly, the overall shape of characters such as the ’Z’ and the ’3’ is better reconstructed. It also appears to enhance stroke reconstruction and improve repositioning, as the model can see during training sections of 3D signals.

Table 3: TEM and TEM-C comparison at label and stroke levels using the Fréchet distance. The number of examples for which a model is better than the other is also computed.

Evaluation	TEM	TEM-C
Label level	0.437	**0.334**
Stroke level	0.097	**0.091**
# Wins	69	**197**

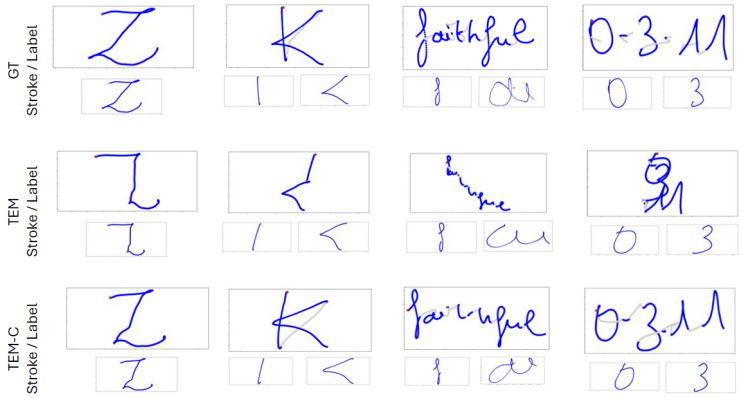


Figure 13: Comparison between the touching expert model (TEM) and its enhancement (TEM-C), on the first line the ground truth at stroke (down) and label (up) level, on the second line the expert model dedicated to touching strokes, on the last line the TEM improved by adding temporal context (TEM-C).

5.4.2. Fine tuning on extra dimension for the hovering expert (HEM-I)

We now evaluate the training of the hovering expert on inclined data, including the extra dimension (discussed in section 4.4). Fine-tuning the HEM model on data acquired from inclined planes has yielded remarkable improvements in prediction accuracy. It can be seen both at the stroke and label level, quantitatively in Table. 4 and

qualitatively in Fig. 14. Our expert model seems to tackle the inherent variability in sensors during hovering parts, that is due to different dynamics between writing and hovering strokes and from having the height dimension varying in the data. Our expert model effectively addresses the inherent variability in sensor data during hovering phases, which arises from the different dynamics between writing and hovering strokes, as well as fluctuations in the height dimension. By training on data from inclined planes to simulate various pencil heights, we enable the model to encounter a broader range of height variability, improving its ability to predict hovering segments. The fine-tuned model (HEM-I) exhibits adaptability to changes in pencil hovering height, resulting in a more robust framework for predicting these segments, as evidenced by the Fréchet distance measurements at the label level. This results in qualitatively better hovering prediction on all examples. Furthermore, improvements at the stroke level can be attributed to a more precise orientation of characters in their reconstructions, stemming from an enhanced understanding of spatial relationships.

Table 4: HEM and HEM-I comparison at label and stroke levels using the Fréchet distance. The number of examples for which a model is better than the other is also computed.

Evaluation	HEM	HEM-I
Label level	0.413	**0.350**
Stroke level	0.129	**0.114**
# Wins	77	**189**

5.4.3. Comparison of model combinations into a mixture-of-experts

Now that each expert has been established, we evaluate the performance of each possible mixture combination (Fig. 15 and Table 5). We introduce some notation for the mixture-of-experts (MOE) :

- The combination touching expert model (TEM) & hovering expert model (HEM) will be noted: MOE;
- The combination touching expert model (TEM) & Hovering expert model fine tuned on inclined data (HEM-I) will be noted: MOE-I;

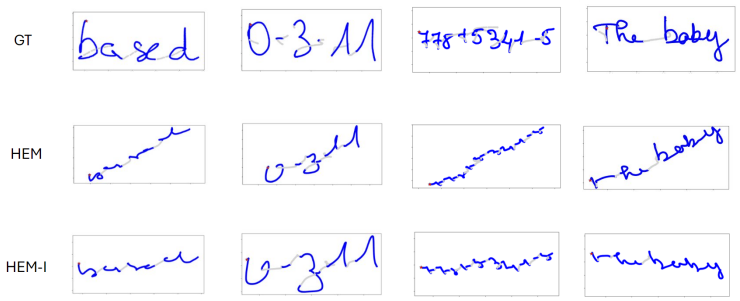


Figure 14: Comparison between the hovering expert model and with a fine-tuning on inclined data, on the first line the ground truth, on the second line the expert model dedicated to hovering (HEM), on the last line the fine-tuned model on inclined data (HEM-I).

- The combination touching expert model with temporal context (TEM-C) & hovering expert model (HEM) will be noted: MOE-C;
- As a reminder, the combination touching expert model with temporal context (TEM-C) & Hovering expert model fine tuned on inclined data (HEM-I) will be noted: MOE-CI.

This evaluation shows us the benefits of different contributions within a mixture of models. The addition of context enables more accurate reconstruction of touching strokes, especially at their extremities. This can be seen visually in the reconstruction of the second stroke of the 'k' or 'a' in 'faithfull'. Note that differences in performance may occur at stroke level for the same expert as the bounding box normalizations are applied at label level. Fine-tuning on 3D data enables better repositioning of hoverings. This is particularly noticeable on long sequences with a lot of hovering, such as the equation where the hovering is much better reconstructed. But also on long hovering sequences, for example the hovering between 'the' and 'baby', which is much better reconstructed. In addition, the combination MOE-CI, which is the combination of the 2 improved experts, is actually the method that shows the best results, as expected. We can see that the reconstructions are better for two reasons: better prediction of hovering and an improvement in stroke. The MOE-CI is more often better than the others, which attests to the benefits of our contributions. Nevertheless, the MOE has

correct performance due to good performances on short mono stroke examples, which are the easiest part to reconstruct.

Table 5: Evaluation of possible mixture of expert on the KIHT-Private dataset. The Fréchet distance is computed to evaluate the models at label and stroke level. Note here the sum of the # Wins line is greater than the number of test labels, since there are common parts in the MOE, so in the event of a tie, the score for the 2 networks is counted.

Evaluation	MOE	MOE-I	MOE-C	MOE-CI
Label level	0.344	0.321	0.333	**0.312**
Stroke level	0.096	0.091	0.096	**0.091**
# Wins	93	86	85	**123**

Figure 15: Comparison between the different combinations of mixture-of-experts, on the first line the ground truth, then from the top to bottom: MOE, MOE-C, MOE-I and MOE-CI.

5.5. Evaluation on the public dataset

We have released the KIHT-public dataset that will serve as a benchmark for futue research works. We evaluate the performance of our mixture-of-experts (MOE-CI) and compare it both to the different expert combinations (MOE, MOE-C, MOE-I) and to the state-of-the art approaches [12, 14]. The results (Table 6) are similar to those obtained on the private dataset. Indeed, the integration of an expert trained at the label level significantly improves performance at the label level compared to the state-of-the-art, and thus improves the processing of repositioning of the reconstructed handwriting

after hovering. The proposals associated with each expert improve the robustness of mixture-of-experts. Whether it's incorporating temporal context that reflects physics and dynamics to enhance the touching expert model, or refining the hovering expert model with 3D labeled samples for improved hovering predictions. These consistent results shows that the public dataset is relevant to be use as benchmark.

Table 6: Fréchet distance computed for our mixture-of-experts and state-of-the-art methods train on the KIHT-Public dataset, and evaluate on the test set. The line "# wins" counts the number of labels where the model obtains the best scores against the others.

Evaluation	[12]	[14]	MOE	MOE-I	MOE-C	MOE-CI
Label level	0.583	0.464	0.336	0.331	0.321	**0.320**
Stroke level	0.127	0.101	0.101	0.100	0.097	**0.096**
# Wins	16	29	68	65	56	**93**

6. Limitations

Dependence on the public dataset may be problematic. Although the dataset has been significantly expanded compared to previous state-of-the-art works, it still cannot cover the spectrum of use cases, particularly in terms of unique handwriting styles and variability in input data. This could potentially limit the model's generalization capabilities across various real-world scenarios.

In addition, while the model demonstrates good ability to generalize across different users, it exhibits significant limitations in handling handwriting from children. Indeed, children have less graphomotor skills, which produces input data that is very different from that of adults, on which the models were trained. In the same way, the model has been trained on tablet writing and the inherent noise associated with paper-based writing significantly hinders its applicability in its current form.

Furthermore, the architecture comprises two dedicated expert models, which requires more memory, limiting the use in an embedded version within the stylus. This presents potential computational challenges in real-time applications, particularly when aiming for deployment in memory-constrained environments while maintaining computational efficiency.

7. Conclusion & perspectives

This paper introduces a novel method for reconstructing handwriting trajectories from kinematic sensor signals embedded in a digital pen, called Digipen designed by STABILO. We present a mixture-of-experts where each model is task-specific. The first expert model predicts touching strokes and processes 2-dimensional signals. The second expert model predicts trajectory repositioning between two touching strokes and processes 3-dimensional inputs due to pen-up movements. We demonstrated the relevance of our proposed improvements for each expert model. We incorporated temporal context that reflects physics and dynamics to enhance the touching expert model. We also fine-tuned the hovering expert model on data acquired on an inclined plane in order to benefit of height dimension variations in training. Our experiments on two datasets demonstrate that our mixture-of-experts outperforms the two main state-of-the-art methods.

In addition, we propose the KIHT-public dataset, a database composed of more than 4k examples. Evaluations have consistently revealed similar conclusions between the KIHT-private dataset and the KIHT-public one, validating the relevance of this new public database. It enables researchers to rigorously test and refine algorithms, facilitating a deeper understanding of the complexities involved in handwriting trajectory reconstruction.

This work is dedicated to the handwriting reconstruction from data written on a tablet using the Digipen. The next challenge is to study the reconstruction of handwriting from data written on paper, which is noisier due to more friction. This task could be solved using transfer learning. In addition, as the Digipen can be used for learning to write in classroom, one should consider domain adaptation methods to transition from adult-acquired data to child-produced data.

8. Acknowledgments

This project is financed by the KIHT French-German bilateral ANR-21-FAI2-0007-01 project and these four partners, IRISA, KIT, Learn & Go and Stabilo. This work was performed using HPC resources from GENCI-IDRIS (Grant 2021-AD011013148)

References

[1] D. Simonnet, N. Girard, E. Anquetil, et al., Evaluation of children cursive handwritten words for e-education, Pattern Recognition Letters (2019).

[2] O. Atilola, S. Valentine, H.-H. Kim, D. Turner, E. McTigue, T. Hammond, J. Linsey, Mechanix: A natural sketch interface tool for teaching truss analysis and free-body diagrams, Artificial Intelligence for Engineering Design, Analysis and Manufacturing 28 (2) (2014) 169–192.

[3] L. Barreto, P. Taele, T. Hammond, A Stylus-Driven Intelligent Tutoring System for Music Education Instruction, 2016, pp. 141–161.

[4] N. Bonneton-Botté, S. Fleury, N. Girard, M. Le Magadou, A. Cherbonnier, M. Renault, E. Anquetil, E. Jamet, Can tablet apps support the learning of handwriting? an investigation of learning outcomes in kindergarten classroom, Computers & Education 151 (2020) 103831.

[5] O. Krichen, S. Corbillé, É. Anquetil, et al., Combination of explicit segmentation with seq2seq recognition for fine analysis of children handwriting, International Journal on Document Analysis and Recognition (IJDAR) (2022).

[6] E. Taranta, A. Vargas, S. Compton, J. Jr, A dynamic pen-based interface for writing and editing complex mathematical expressions with math boxes, ACM Transactions on Interactive Intelligent Systems (TiiS) 6 (2016) 13. doi:10.1145/2946795.

[7] M. Wilhelm, D. Krakowczyk, S. Albayrak, Perisense: Ring-based multi-finger gesture interaction utilizing capacitive proximity sensing, Sensors (2020). doi:10.3390/s20143990.

[8] V. J. Pablo, B. L. L. Isabel, V. C. Ángel Leonardo, B. M. E., Hand gesture recognition using emg-imu signals and deep q-networks, Sensors 22 (24) (2022). doi:10.3390/s22249613.

[9] H. Choi, H. Jeon, D. Noh, T. Kim, D. Lee, Hand-guiding gesture-based telemanipulation with the gesture mode classification and state estimation using wearable imu sensors, Mathematics 11 (16) (2023).

[10] S. Derrode, H. Li, L. Benyoussef, et al., Unsupervised pedestrian trajectory reconstruction from IMU sensors, in: TAIMA 2018: Traitement et Analyse de l'Information Méthodes et Applications, Tunisia, 2018.

[11] E. Digo, L. Gastaldi, M. Antonelli, S. Pastorelli, A. Cereatti, M. Caruso, Real-time estimation of upper limbs kinematics with imus during typical industrial gestures, Procedia Computer Science 200 (2022) 1041–1047.

[12] M. Wehbi, D. Luge, T. Hamann, et al., Surface-free multi-stroke trajectory reconstruction and word recognition using an imu-enhanced digital pen, Sensors (2022).

[13] F. Ott, D. Rügamer, L. Heublein, et al., Joint classification and trajectory regression of online handwriting using a multi-task learning approach, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 266–276.

[14] S. Wassim, F. Imbert, S. Yann, R. Tavenard, É. Anquetil, Online handwriting trajectory reconstruction from kinematic sensors using temporal convolutional network, International Journal on Document Analysis and Recognition (IJDAR) (2023) 1–14.

[15] T. Miyagawa, Y. Yonezawa, K. Itoh, M. Hashimoto, Handwritten pattern reproduction using pen acceleration and angular velocity, IEICE Trans. (2000).

[16] T.-Y. Pan, C.-H. Kuo, M.-C. Hu, A noise reduction method for imu and its application on handwriting trajectory reconstruction, in: 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2016, pp. 1–6. doi:10.1109/ICMEW.2016.7574685.

[17] T. Deselaers, D. Keysers, J. Hosang, H. A. Rowley, Gyropen: Gyroscopes for pen-input with mobile phones, IEEE Transactions on Human-Machine Systems 45 (2) (2015) 263–271. doi:10.1109/THMS.2014.2365723.

[18] T. Harbaum, A. Serdyuk, F. Kreß, T. Hamann, J. Barth, P. Kämpf, F. Imbert, Y. Soullard, R. Tavenard, E. Anquetil, J. Delahaie, KIHT: Kaligo-based Intelligent Handwriting Teacher, in: DATE 2024, Valencia, Spain, 2024.

[19] Z. Chen, D. Yang, J. Liang, et al., Complex handwriting trajectory recovery: Evaluation metrics and algorithm, in: Proceedings of the Asian Conference on Computer Vision, 2022, pp. 1060–1076.

[20] E. Mohamed Moussa, T. Lelore, H. Mouchère, Point to segment distance DTW for online handwriting signals matching, in: ICPRAM 2023, SCITEPRESS - Science and Technology Publications, Lisbonne, Portugal, 2023, pp. 850–855. doi:10.5220/0011672600003411.

[21] Y. Bu, L. Xie, Y. Yin, et al., Handwriting-assistant: Reconstructing continuous strokes with millimeter-level accuracy via attachable inertial sensors, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (2021).

[22] Y. Wang, J. Xu, Y. Zhao, Wavelet encoding network for inertial signal enhancement via feature supervision, IEEE Transactions on Industrial Informatics PP (2024) 1–11. doi:10.1109/TII.2024.3431018.

[23] Y. Wang, Y. Zhao, Scale and direction guided gan for inertial sensor signal enhancement, in: K. Larson (Ed.), Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24, International Joint Conferences on Artificial Intelligence Organization, 2024, pp. 5126–5134, main Track. doi:10.24963/ijcai.2024/567.
URL https://doi.org/10.24963/ijcai.2024/567

[24] X. Chen, Z. Li, Y. Yang, L. Qi, R. Ke, High-resolution vehicle trajectory extrac-

tion and denoising from aerial videos, IEEE Transactions on Intelligent Transportation Systems PP (2020) 1–13. doi:10.1109/TITS.2020.3003782.

[25] X. Chen, J. Ling, Y. Yang, H. Zheng, P. Xiong, O. Postolache, Y. Xiong, Ship trajectory reconstruction from ais sensory data via data quality control and prediction, Mathematical Problems in Engineering 2020 (2020) 1–9. doi:10.1155/2020/7191296.

[26] Y. Liu, K. Huang, X. Song, et al., Maghacker: Eavesdropping on stylus pen writing via magnetic sensing from commodity mobile devices, in: Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services, MobiSys '20, 2020.

[27] A. Serdyuk, F. Kreß, M. Hiegle, T. Harbaum, J. Becker, F. Imbert, Y. Soullard, R. Tavenard, E. Anquetil, J. Barth, P. Kämpf, Towards the on-device Handwriting Trajectory Reconstruction of the Sensor Enhanced Pen, in: IEEE 9th World Forum on Internet of Things, 2023.

[28] F. Ott, D. Rügamer, L. Heublein, et al., Benchmarking online sequence-to-sequence and character-based handwriting recognition from imu-enhanced pens, arXiv preprint arXiv:2202.07036 (2022).

[29] F. Ott, M. Wehbi, T. Hamann, et al., The onhw dataset: Online handwriting recognition from imu-enhanced ballpoint pens with machine learning, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4 (3) (2020) 1–20.

[30] L. Wegmeth, A. Hoelzemann, K. Van Laerhoven, Detecting handwritten mathematical terms with sensor based data, arXiv preprint arXiv:2109.05594 (2021).

[31] F. Kreß, A. Serdyuk, T. Hotfilter, et al., Hardware-aware workload distribution for ai-based online handwriting recognition in a sensor pen, in: 2022 11th Mediterranean Conference on Embedded Computing (MECO), IEEE, 2022, pp. 1–4.

[32] A. Klaß, S. M. Lorenz, M. Lauer-Schmaltz, et al., Uncertainty-aware evaluation of time-series classification for online handwriting recognition with domain shift, arXiv preprint arXiv:2206.08640 (2022).

[33] F. Ott, D. Rügamer, L. Heublein, et al., Domain adaptation for time-series classification to mitigate covariate shift, in: the 30th ACM International Conference on Multimedia, 2022, pp. 5934–5943.

[34] H. Azimi, S. Chang, J. Gold, et al., Improving accuracy and explainability of online handwriting recognition, preprint arXiv:2209.09102 (2022).

[35] Y. Wang, Y. Zhao, Handwriting recognition under natural writing habits based on a low-cost inertial sensor, IEEE Sensors Journal 24 (1) (2024) 995–1005. doi:10.1109/JSEN.2023.3331011.

[36] H. T. Nguyen, T. Nakamura, C. T. Nguyen, et al., Online trajectory recovery from offline handwritten japanese kanji characters of multiple strokes, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 8320–8327.

[37] H. Huang, D. Yang, G. Dai, et al., Agtgan: Unpaired image translation for photographic ancient character generation, in: the 30th ACM International Conference on Multimedia, 2022, pp. 5456–5467.

Florent Imbert is a Ph.D student at INSA. He hold a master's degree in applied mathematics from La Rochelle University. His Ph.D. is being carried out at IRISA and focuses on online handwriting trajectories reconstruction from kinematic sensors.

Eric Anquetil is a full professor at INSA Rennes. He leads the research on handwriting, gesture and drawing analysis and recognition at the IRISA laboratory. He is in charge of the "Innovation and Entrepreneurship" mission at INSA Rennes. His expertise is in human-machine interactivity through handwriting input and gesture commands.

Yann Soullard is an associate professor at Rennes 2 University (France) and a member of the IRISA computer science research institute. He holds a PhD degree from the Pierre and Marie Curie university of Paris, obtained in 2013. His research interests include deep learning, document analysis and handwriting recognition.

Romain Tavenard is a full professor at University of Rennes 2 since September 2022, following nine years as an assistant professor. He conducts research at IRISA, with a focus on machine learning and indexing for time series data, specifically for environmental time series.