



# Automatic monitoring of steel strip positioning error based on semantic segmentation

Aline de Faria Lemos<sup>1</sup> · Leonardo Adolpho Rodrigues da Silva<sup>2</sup> · Balázs Vince Nagy<sup>1</sup>

Received: 22 April 2020 / Accepted: 30 July 2020 / Published online: 1 September 2020  
© The Author(s) 2020, corrected publication 2020

## Abstract

The misalignment of steel strips in relation to the roller table centerline still is an impairment for the rolling mill production lines. Nowadays, the strip position correction remains largely in the purview of human analysis, in which the strip steering is traditionally a semi-manual operation. Automating the alignment process could reduce the maintenance costs, damage to the plant, and prevent material losses. The first step into the automatization is to determine the strip position and its referred error. This study presents a method that employs semantic segmentation based on convolution neural networks to estimate steel strips positioning error from images of the process. Additionally, the system mitigates the influences of mechanical vibration on the images. The system performance was assessed by standard semantic segmentation evaluation metrics and in comparison with the dataset ground truth. The results showed that 97% of the estimated positioning errors are within a 2-pixel margin. The method demonstrated to be a robust real-time solution as the networks were trained from a set of low-resolution images acquired in a complex environment.

**Keywords** Steckel mill · Steel strips positioning error · Semantic segmentation · Convolution neural network · Hot strips

## 1 Introduction

Steel strips are manufactured from cast slabs, which undergo several times between a pair of work rolls with decreasing gaps until the achievement of the intended thickness reduction [20, 21, 39, 40]. In Steckel mill lines, during the hot rolling process, the strips are driven by the roller table in the rolling direction. However, the rolling procedure is susceptible to impel the strips perpendicularly to this direction, which could induce misalignment. For instance, the unaligned strips are prone to collide

with the side guides and mill structure [4, 11]. As the 20-ton strips are rolled at 10 m/s, the collisions are an impairment for the production lines, provoking material losses and damaging the mill structure and equipment. Annually, the material loss due to collisions and equipment failure expenses are about one million euros [11].

Traditionally, the process of correcting the alignment of the strips is semi-manual. In this process, a human operator observes the strip position through real-time images of the process, which are acquired from analog cameras settled over the mill structure. From the content of the images, the operator deduces whether the strip presents deviation from the roller table centerline. Per this information, the operator attempts to recenter the strip by a manual command, which adjusts the gap between the extremities of the work rolls asymmetrically and steers the strip on the direction of the larger gap difference [4]. The realignment correction procedure requires a reaction time that might exceed the human capability, becoming susceptible to failure due to the high longitudinal speed of the strips. In addition, a manual command can lead to inadequate control, as it is an imprecise tool [9, 11].

Automating the steel strips alignment process in Steckel mill lines could reduce the maintenance costs, damage to the plant, and prevent material losses [4, 11]. The first step through the process automatization is to determine the strip

✉ Aline de Faria Lemos  
alinefaria@mogi.bme.hu

Leonardo Adolpho Rodrigues da Silva  
leonardo@ufsj.edu.br

Balázs Vince Nagy  
nagyb@mogi.bme.hu

<sup>1</sup> Department of Mechatronics, Optics and Mechanical Engineering Informatics, Budapest University of Technology and Economics, Budapest, Hungary

<sup>2</sup> Department of Telecommunications and Mechatronics Engineering, Federal University of São João del-Rei, Ouro Branco, Brazil

positioning error, which could be accessed from the images of the process.

The available literature presents a considerable number of studies with solutions for metalworking. These include the detection of defects in metal casting [5, 10], recognition of slab identification [22], prediction of mechanical properties [38], bearing fault diagnosis [14, 19, 29, 35, 41], and steel defect identification [25, 28, 31]. On the other hand, there are only a few works that apply image processing to measure the strip position in a rolling mill process. From the best of the knowledge of the authors, only two studies could be found. The first applies traditional image processing applying Bezier curves to measure the strip centerline in a rolling process [4]. This study was performed in a dataset with limited size. The second study applies morphological operations in processed images to calculate steel strip positioning error [9]. However, the work lacks validation from a ground truth set.

In this paper, a novel method to access the strip positioning error is presented. The system employs semantic segmentation based on convolution neural network (CNN) to extract the strip portion of the images of the process, infers the strip position, and estimates the positioning error in relation to setpoint images. Moreover, an analogous approach is performed to attenuate the influences of the camera mechanical vibration to the images. It consists of applying a CNN-based semantic segmentation method to identify the position of a static mill component that appears on the images. This position is latter selected as the reference point of the strip relative position.

The system performance was assessed by standard semantic segmentation evaluation metrics and in comparison with the positioning error signal derived from the dataset ground truth. The method proved to be a robust real-time solution as the networks were trained from a set of low-resolution images acquired in a complex environment, containing steam and variable luminance.

The remainder of this paper is divided into the following sections: “Theoretical basis,” “Methods,” “Results and discussion,” and “Conclusions”.

## 2 Theoretical basis

### 2.1 Process description

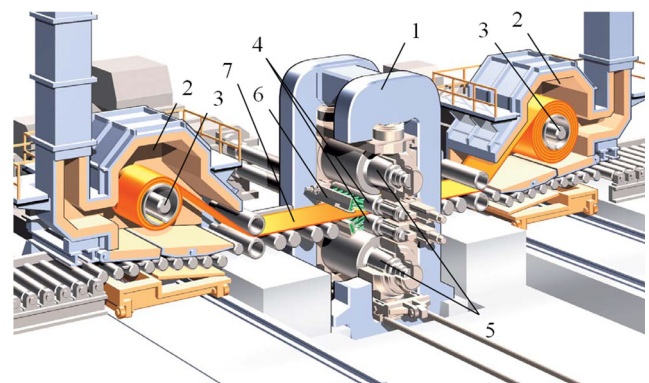
The rolling mill process consists of the thickness reduction of steel slabs from successively passing it through a pair of work rolls with a decreasing gap. The Steckel mill process differs from the traditional rolling by the presence of coiler furnaces in either side of the rolling stand, as illustrated in Fig. 1, which reheat the strips while wounding upon their drums. This process allows the product to reach lengths

approximately equal to 600 m [11, 20]. The carbon steel heating produces an oxide layer over the strip, which is removed before each pass of the strip through the work rolls by high-pressure water, in a process denominated descaler, originating heavy steam.

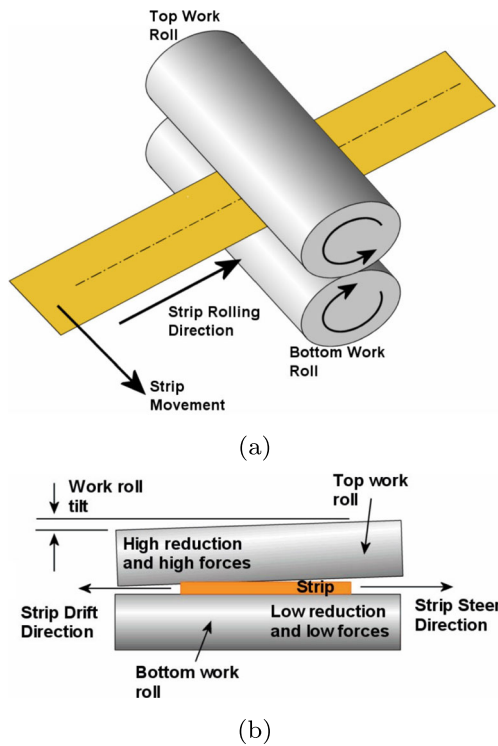
During a few steps of the process, the side guides align the strip in relation to the roller table centerline before the rolling operation. From this position, the strips are moved by the roller table in the rolling direction. Moreover, the rolling induces the strips to move perpendicularly to this direction, according to the illustration present in Fig. 2a. This fact tends to cause a strip misalignment, which could lead to collisions and, consequently, process losses [11]. Nowadays, the strip position correction remains largely in the purview of human analysis [4]. The realignment process is executed by an operator, which judges the strip misalignment from real-time images of the process. The operator attempts to compensate for the undesirable direction of the movement tilting the work rolls with a manual command. This procedure creates an asymmetrical gap between the work rolls, inducing higher reduction forces on the side of the smaller gap and lower forces on the opposite side. Therefore, the wedging effect will steer the strip in the direction of the larger gap, which corresponds to the direction of the desired position [11], as shown by Fig. 2b.

### 2.2 Semantic segmentation

Semantic segmentation is a pixel-wise categorization, which gathers pixels belonging to the same class [24, 28, 30]. Regarding digital image processing, the method is best applied as an emulator for human pattern identification [28]. Compared to the traditional image segmentation, semantic segmentation based on convolution neural network has demonstrated considerable advantages [28] and has been applied to many tasks such as medical applications [2,



**Fig. 1** Steckel reversible rolling mill with coiler furnaces. 1: rolling stand, 2: coiler furnace, 3: winding drum of the coiler furnace, 4: work roll, 5: backup roll, 6: descaler jets, 7: steel strip. Source [1, 20]



**Fig. 2** Strip misalignment during rolling and correction through roll tilting [11]. **a** Strip movement directions. **b** Effects of roll tilting. The procedure creates an asymmetrical gap between work rolls, in which the high reduction forces steer the strip in the direction of the larger gap

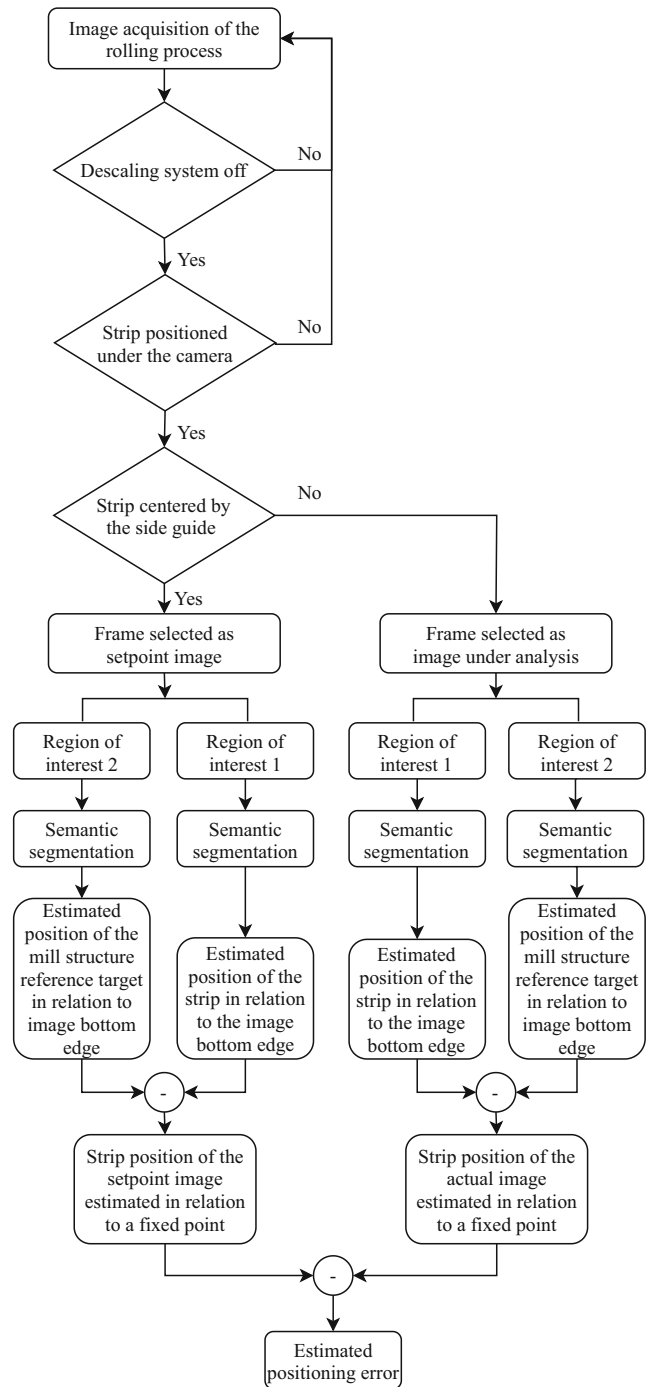
26, 27, 42], in autonomous driving [6, 8, 12, 32], object detection [13, 37], and pose estimation system [33], to name a few. The semantic segmentation architecture usually consists of an encoder-decoder task [3, 15, 23]. The first part, composed of convolution and max pooling operations, extracts high-level features by mapping the input to a lower dimension representation [23, 36]. On the other hand, the architecture of the decoder, commonly composed of transposed convolutions and up-pooling layers, expands the high-level features, recovering the feature map size compatible with the input layer size [28].

### 3 Methods

This work adopted supervised learning to estimate the positioning error of steel strips in a Steckel mill line. The method employs hybrid semantic segmentation to estimate the strip position through images of the process and calculates their positioning error in contrast to values derived from setpoint (or reference) images. Additionally, the system mitigates the influences of mechanical vibration on the process videos. A concise explanation of the adopted

methodology is presented in the flowchart of Fig. 3, and further clarification is exposed in the section remainder.

The dataset is composed of RGB images, which were acquired from an analog camera installed over the mill stand entrance, on the operator side, with a sample rate of 30 fps. From the acquired images, the algorithm gathered the images of interest according to the activation command of the descaler and strip tracking signals. During the descaler



**Fig. 3** System flow chart

process, the heavy steam content present on the region between camera and strip makes the image manipulation unfeasible. Also, during part of the acquisition time, the strip is not positioned on the field of view of the camera. Therefore, the images captured in both circumstances will not be processed by the positioning error estimation algorithm.

Succeeding the dataset selection, the centralization command signal of the side guides is used to categorize the images of interest into position setpoint images or images under analysis. When the signal is active, the images are classified into setpoint images, since the side guides ensure the strip centralization, as aforementioned in Section 2.1. The system stores the reference image and estimates the strip position of both categories. The positioning error is then calculated by comparing the actual image strip position and the position of the last detected setpoint image.

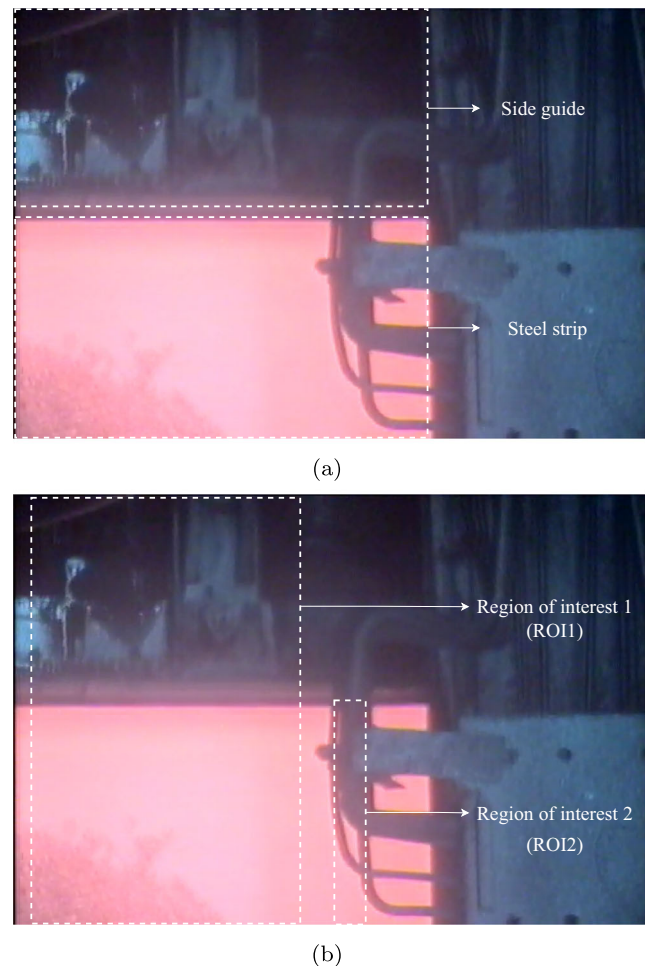
### 3.1 Regions of interest

An example of the images of the process used as the system input is presented in Fig. 4a. As highlighted on the figure, the images contain a portion of the strip aligned to the image bottom edge, side guide elements, and parts of the mill structure. Therefore, to reduce the image complexity, a region of interest for strip position estimation (ROI1) was elected, lowering the number of mill components present on the image.

The impacts of the rolling process cause unwanted camera vibration, which is unavoidable as the camera is placed 6 m above the strip. This fact can be observed in the image present in Fig. 5a, which shows parts of the mill structure of two consecutive frames. From this figure, it is perceptible that the distance between the mill structures and the horizontal dashed line varies in a considerable amount. Empirical observations revealed that the mill structures adjacent to the strip present an irrelevant relative movement in relation to the strip. Hence, to avoid interference from the camera vibration effects on the estimated values, the mill structure parts visible on the images of the process were used as a strip position reference. These structures are present in the region of interest 2 (ROI2), indicated in Fig. 4b. In ROI2, the mill parts delimit a polygon, which centroid is used as the mentioned position reference. This polygon and its centroid are highlighted in Fig. 5b.

### 3.2 Labeling

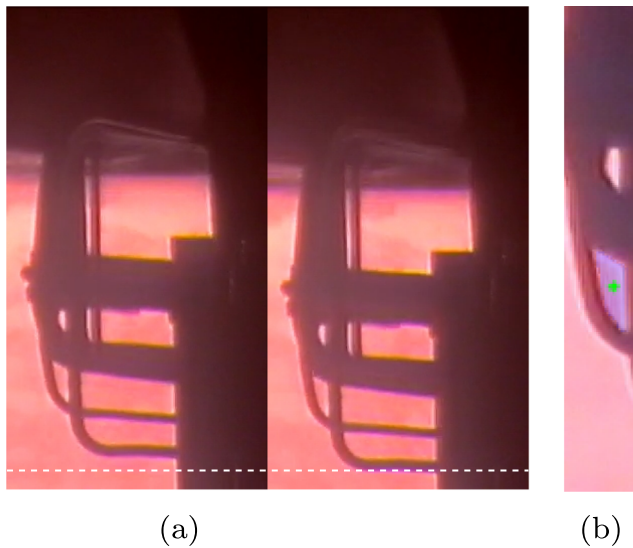
The ground truth labeling of each region of interest was created by manual annotation utilizing the Image Labeler Matlab app. Pixels belonging to the reference polygon or strip portion were assigned to the intensity value 1, and pixels belonging to the background were assigned to value



**Fig. 4** Example of the images of the process used as the system input. **a** Location of the side guides and steel strip on the image. **b** Elected regions of interest used to estimate the strip position (ROI1) and to reduce the effects of the camera vibration (ROI2)

2. In the total, 1390 labels were created for each region of interest, as the selected dataset comprises  $1390\ 704 \times 480$ -px RGB images.

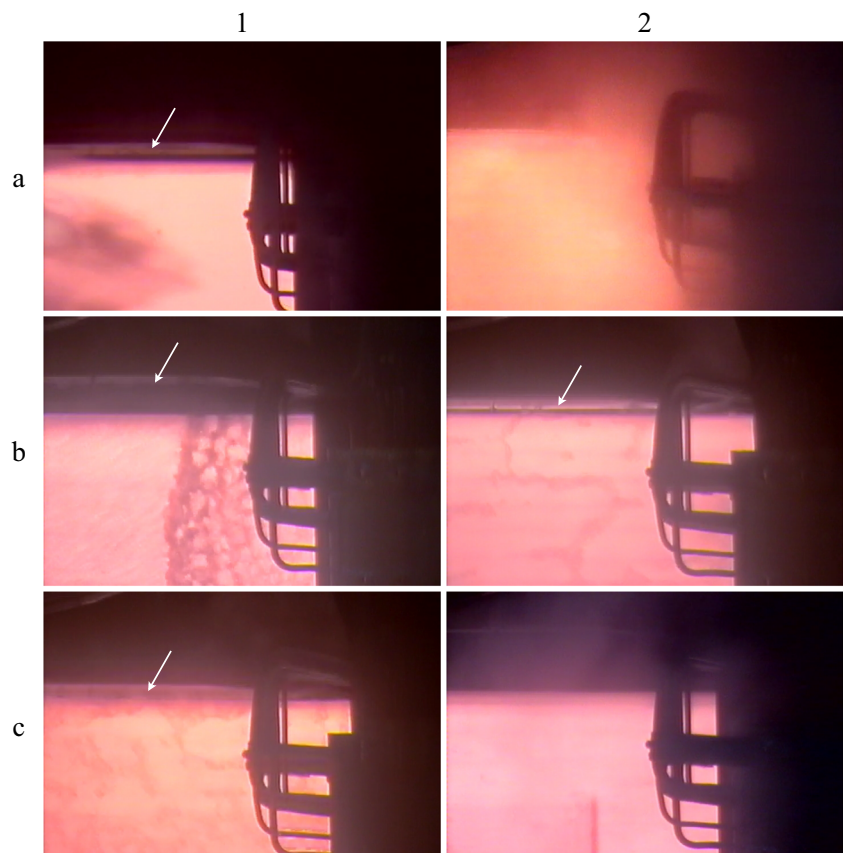
The images were acquired in a complex environment. The remaining elements from the descaler process, such as water over the strip and steam content on the strip location and surroundings, compromise the image quality. The water creates unpredictable patterns over the strip, as can be perceived in Fig. 6-2b, 2c, 1b, and 1c. On the other hand, the steam content blurs the acquired images. Figure 6-2a, 2c, and 1a show some of these blurring particularities. Another complication is the strip incandescence, resulting from the strip high temperature, which reflects over the side guide structures present on ROI1. These structures mirror the strip color and could be easily mistaken as strip portions, similar to the effects indicated in Fig. 6-2b, 1a, 1b, and 1c by white arrows. The labeling process was handled carefully by considering these occasions, to avoid misclassifications.



**Fig. 5** Example of camera vibration. **a** Distance from the mill structures to the image bottom edge and the horizontal dashed line resulted from camera vibration; **b** Centroid of the polygon used as the camera reference position of the strips on the images instead of the image bottom edge

In cases that portions of the strip were covered by steam, the labeling considers that the strip location is parallel to the image bottom.

**Fig. 6** Example of complex elements on the acquired images. Images 1a, 2a, and 2c are blurred due to the steam content, and remaining water from the descaler process can be perceived in images 1b, 1c, 2b, and 2c as unpredictable patterns over the strip. The reflections of the strip incandescence over the side guide structures are indicated by white arrows

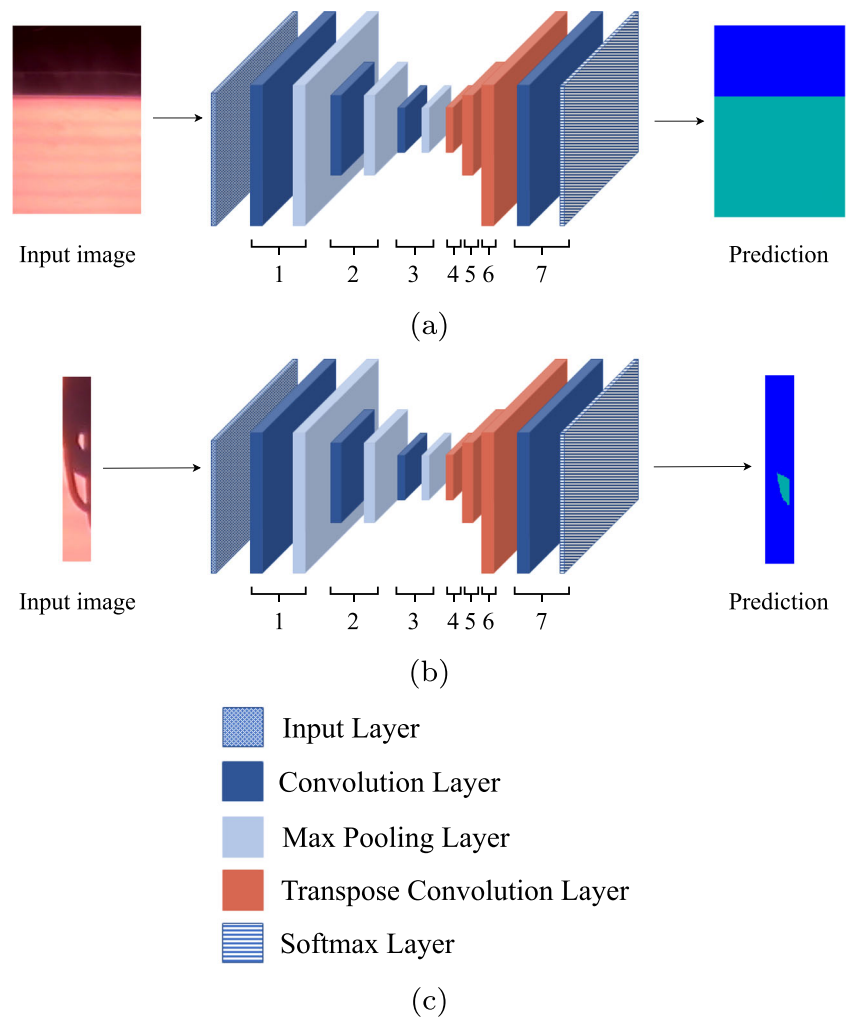


### 3.3 Semantic segmentation

In the present work, two semantic segmentation approaches based on CNN operations were applied to detect, independently, the strip portion present in the ROI1 and the polygon delimited by the mill components present in ROI2. Three architecture configurations were investigated for each region of interest (ROI). The architectures differ from each other by the number of encoder/decoder layers, varying between 1, 2, or 3 pairs. Figure 7 illustrates the largest network in terms of the number of layers (three encoders and three decoders) applied to each ROI. The encoder part consists of downsampling layers, which include convolution and max pooling operations. In contrast, the decoder architecture comprises upsampling layers, which consists of transposed convolutions. After each convolution and transposed convolution layer, the Rectified Linear Unit (ReLU) activation function was applied. Adam optimizer and a learning rate of 0.001 were selected as optimization parameters. Moreover, the influence of the number of filters in each operation was also ascertained. The number of filters could hold values from the set {2, 4, 8, 16, 32, 64}. Hence, eighteen architectures were explored altogether for each ROI.

The encoder-decoder architecture is followed by a pixel classification layer, which enables a pixel-level classification and is composed of convolution and a Softmax

**Fig. 7** Schematic illustration of one of the semantic segmentation architectures designed for **a** ROI1 and **b** ROI2. 1: Downsampling layer 1, 2: Downsampling layer 2, 3: Downsampling layer 3, 4: Upsampling layer 3, 5: Upsampling layer 2, 6: Upsampling layer 1, 7: Pixel Classification Layer. **c** Legend



layer. This last convolution operation combines the input feature maps down to the number of classes. Hence, this layer is configured with kernel size and the number of classes equal to 1 and 2, respectively. The further layer configurations are presented in Table 1, in which the layer nomenclatures refer to those introduced in Fig. 7.

Concerning the dataset, it was initially composed of 1390 images, which was split into training and test datasets on the proportion of 1112 to 278 images, respectively.

### 3.4 Positioning error estimation

The CNN predictions of the ROI1 were refined via morphological operations and outlier exclusion to adjust misclassifications induced by the presence of complex elements in the image. Details about these elements are mentioned in Section 3.2 and exemplified in Fig. 6. The probabilistic mask of the strip, obtained from the segmentation, was binarized. From the image, the connected components were identified and deleted, except for the larger connected component. This operation keeps the strip

area and eliminates smaller and disconnected elements erroneously classified as part of the strip portion, such as steam content and mill components that reflect the strip incandescence. Afterward, the possible holes in the strip area, mostly caused by water over the strip and fluctuation in the incandescence intensity, were filled by a flood-fill operation. From the resulting binary image, the pixel locations of the top edge of the strip portion were employed to estimate the strip position. Then, an outlier removal with a threshold between 40 and 60% was utilized to prevent interference from possible irregular edges. The strip position in relation to the image bottom edge ( $Y_{ImStrip}$ , for the actual image, and  $Y_{StpStrip}$ , for the setpoint image) was determined as the average of the remaining values.

Similarly, the improvement of the ROI2 polygon portion predicted by semantic segmentation was performed via morphological operations, which included the application of flood-fill, erode, and dilate. The flood-fill operation fills holes in the polygon prediction originated by misclassification. The erode and dilate functions separate mill components, other than the desired polygon, into smaller and disconnected components. Thus, a greater connected

**Table 1** Models parametrization

Model ID	Architecture	Number of filters	Kernel size	Downsampling			Upsampling		
				Layer 1 Pooling / stride	Layer 2 Pooling / stride	Layer 3 Pooling / stride	Layer 3 Upsampling factor	Layer 2 Upsampling factor	Layer 1 Upsampling factor
1	1	2	3x3	2/2	–	–	–	–	2
2	1	4	3x3	2/2	–	–	–	–	2
3	1	8	3x3	2/2	–	–	–	–	2
4	1	16	3x3	2/2	–	–	–	–	2
5	1	32	3x3	2/2	–	–	–	–	2
6	1	64	3x3	2/2	–	–	–	–	2
7	2	2	3x3	2/2	2/2	–	–	2	2
8	2	4	3x3	2/2	2/2	–	–	2	2
9	2	8	3x3	2/2	2/2	–	–	2	2
10	2	16	3x3	2/2	2/2	–	–	2	2
11	2	32	3x3	2/2	2/2	–	–	2	2
12	2	64	3x3	2/2	2/2	–	–	2	2
13	3	2	3x3	2/2	2/2	3/3	3	2	2
14	3	4	3x3	2/2	2/2	3/3	3	2	2
15	3	8	3x3	2/2	2/2	3/3	3	2	2
16	3	16	3x3	2/2	2/2	3/3	3	2	2
17	3	32	3x3	2/2	2/2	3/3	3	2	2
18	3	64	3x3	2/2	2/2	3/3	3	2	2

component selection keeps the polygon area and exclude the disconnected elements. The mill structure reference position in relation to the image bottom edge ( $Y_{Imc}$ , for the actual image, and  $Y_{Stpc}$ , for the setpoint image) was calculated from the centroid of the remaining component.

The illustrations presented in Fig. 8 show an example of an image under analysis, Fig. 8a, and a setpoint image, Fig. 8b. Besides, the figures also present the reference systems used to derive the strip positions and further strip positioning error. The global coordinate system ( $XOY$ ) is used to derive the strip position in relation to the image bottom edge ( $Y_{ImStrip}$ , for the actual image, and  $Y_{StpStrip}$ , for the setpoint image). On the other hand, the local coordinate system ( $xcy$ ) aims to mitigate the camera vibration influences over the strip position estimation by providing a static reference position in relation to the process environment. This coordinate system has its origin at the centroid of the polygon composed by the mill components ( $c$ ), and it is located apart from the X-axis in  $Y_{Imc}$ , for the actual image, and  $Y_{Stpc}$ , for the setpoint image. The strip position relative to this coordinate system can be calculated by Eq. 1 (actual image) and Eq. 2 (setpoint image). The strip positioning error is given by the difference between both values (Eq. 3).

- Strip position of the actual image relative to the local coordinate system (mill structure reference point)

$$y_{ImStrip} = Y_{ImStrip} - Y_{Imc} \tag{1}$$

- Strip position of the setpoint image relative to the local coordinate system (mill structure reference point)

$$y_{StpStrip} = Y_{StpStrip} - Y_{Stpc} \tag{2}$$

- Strip positioning error

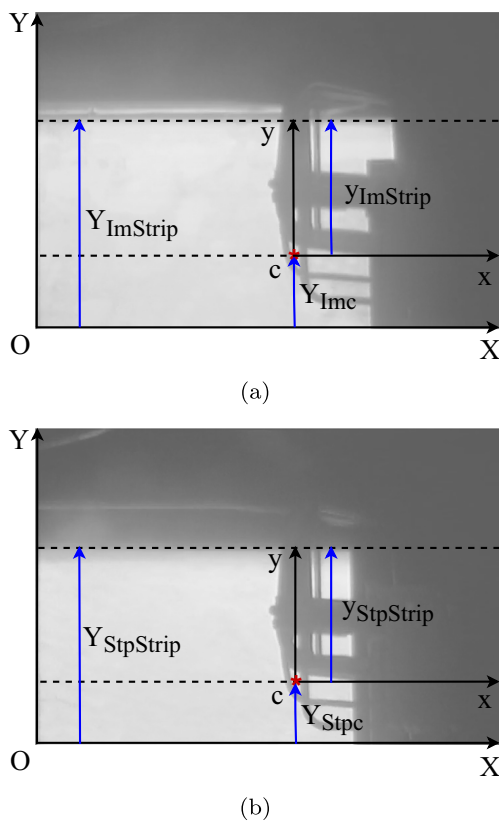
$$y_{error} = y_{ImStrip} - y_{StpStrip} \tag{3}$$

### 3.5 Representation of the positioning error in physical units

On the periods when the strip is not positioned under the camera one of the roller table rolls is visible. The diameter of the roll can be measured on the image and is equal to 140 px, while its physical dimension is 400 mm. Therefore, the resolution of the images is equal to 2.9 mm/px.

### 3.6 Performance evaluation

The proposed method is evaluated by comparing the estimated values of strip position, y-coordinate of the mill



**Fig. 8** Schematic representation of the global and local coordinate systems used to estimate the strip position relative to the mill components of the **a** actual image and **b** setpoint image. In both figures,  $XOY$  is the global coordinate system and  $xcy$  is the local coordinate system

reference position, and positioning error to the expected values, calculated from the ground truth images, by the mean absolute error (MAE) and the standard deviation (STD). The computational burden was also analyzed through frame rate analysis to assess the real-time viability of the application, and the execution was carried out on a NVIDIA GeForce GTX1080 GPU. Also, each architecture evaluation was performed on the test sets by common metrics, applied to evaluate semantic segmentation based on convolution neural networks. These metrics are recall (Eq. 4), Jaccard index (Eq. 5), F1 score (Eq. 6), and specificity (Eq. 7), and they are determined from the pixel predictions of the segmented mask, which are the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [7, 24, 28, 33, 34]

- Recall

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

- Intersection-over-Union (IoU) or Jaccard index

$$IoU = \frac{TP}{TP + FP + FN} \quad (5)$$

- F1 score (F1)

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (6)$$

- True negative rate (TNR) or specificity

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

## 4 Results and discussion

This paper presents an accurate and automatic strip positioning error estimator for a rolling mill process. The system as a whole is composed of a couple of distinct steps, as discussed in Section 3. As a matter of fact, the performance of each step can be evaluated separately. In this section, the obtained results are discussed in the following order: (i) steel strip edge and mill structure location estimation and (ii) steel strip misalignment evaluation. The results present on Section 4.1 are referring to the test set, while analyses of Section 4.2 are based on both test and training sets.

### 4.1 Steel strip edge and mill structure location estimation

At the base of the developed system lies a semantic segmentation neural network. The purpose of this step is to classify each pixel in the current frame as an object of interest (steel strip or mill structure component) or as background. By performing this classification, it is possible to reject most of the irrelevant information present in the image. The results attained by the CNN model does not hold much meaning for the final application on its own. Because of that, the results of position estimation are presented together. Results for the first region of interest are shown in Table 2, while Table 3 shows results for the second region. Both regions are judged by the same metrics. Moreover, the first four results columns (recall, IoU,  $F_1$  score, and specificity) refer to the segmentation step. In contrast, the latter two (MAE and STD) relate to the position estimation errors of the components.

The identification of the strip by the convolution network is performed with very high accuracy. For all types of architectures and numbers of convolutional filters, all four considered metrics exhibit values above 99%. Since the effectiveness of every model is virtually the same, the model choice should be based on the remaining criteria. Before turn to model accuracy, the computational burden of the system must be assessed, since real, or at least quasi-real, execution time is sought. The frame rate of the camera feed is 30 fps. Therefore, it is a desirable model with a superior execution rate, as this rate refers only to this intermediate



**Table 2** Performance evaluation of ROI1

Model ID	Architecture	No. of filters	Recall [%]	IoU [%]	F1 [%]	TNR [%]	Frame rate [fps]	MAE [mm]	STD [mm]
1	1	2	99.7	99.1	99.6	99.0	53.7	5.0	5.2
2	1	4	99.8	99.2	99.6	99.0	51.1	4.3	4.8
3	1	8	99.8	99.4	99.7	99.3	40.8	3.4	4.0
4	1	16	99.9	99.3	99.7	99.1	26.1	3.7	4.3
5	1	32	99.8	99.4	99.7	99.3	7.8	3.7	4.1
6	1	64	99.8	99.3	99.7	99.2	1.6	3.6	4.2
7	2	2	99.8	99.1	99.6	98.8	42.5	4.8	4.9
8	2	4	99.8	99.3	99.7	99.1	52.5	3.8	4.7
9	2	8	99.8	99.4	99.7	99.2	40.0	3.7	3.8
10	2	16	99.9	99.3	99.6	98.9	24.5	4.3	4.4
11	2	32	99.9	99.3	99.7	99.1	6.1	3.7	3.9
12	2	64	99.9	99.4	99.7	99.1	1.6	3.4	3.9
13	3	2	99.8	99.2	99.6	99.0	41.2	5.2	4.6
14	3	4	99.8	99.4	99.7	99.3	47.1	3.7	4.2
15	3	8	99.9	99.4	99.7	99.1	38.5	3.9	3.6
16	3	16	99.9	99.4	99.7	99.0	23.9	4.2	3.4
17	3	32	99.9	99.4	99.7	99.2	5.4	3.8	3.3
18	3	64	99.9	99.5	99.7	99.3	1.6	3.4	3.3

step. The final execution frequency will be presented later in this section. Taking this and the achieved strip estimation accuracy, the elected model configuration, which better balances both objectives, is Model 3. With architecture type 1 (one encoder/decoder pair) and 8 convolutional filters in

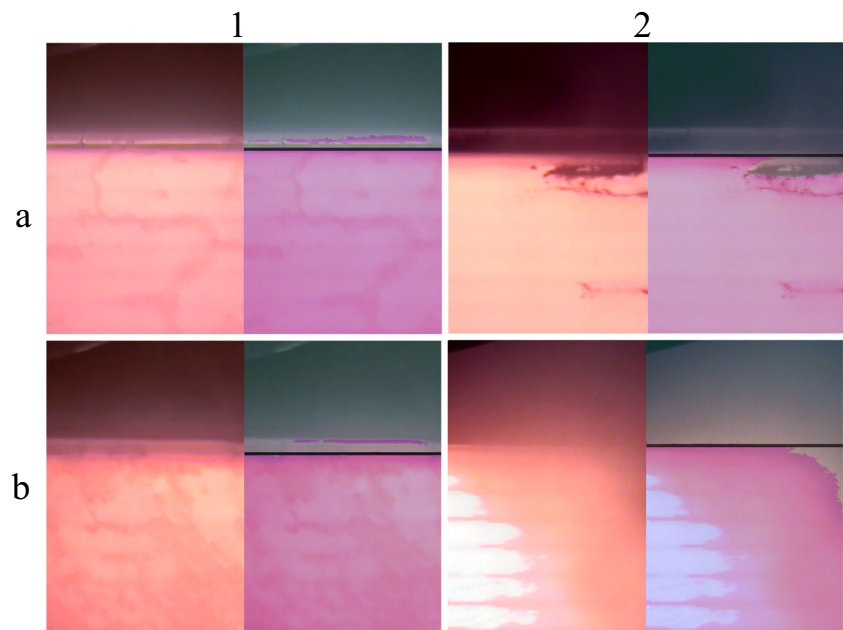
each layer, the model achieved mean absolute error of 3.4 ( $\pm 4.0$ ) mm and frame rate of 40.8 fps.

For the identification of the mill structure, the segmentation process yield very poor results. The main reasons behind this inefficiency are twofold: there is a severe

**Table 3** Performance evaluation of ROI2

Model ID	Architecture	No. of filters	Recall [%]	IoU [%]	F1 [%]	TNR [%]	Frame rate [fps]	MAE [mm]	STD [mm]
1	1	2	13.5	13.4	23.7	99.8	38.9	157.3	18.7
2	1	4	49.7	48.7	65.5	99.8	121.2	8.1	8.9
3	1	8	57.1	56.6	72.3	99.9	126.1	6.4	9.9
4	1	16	68.0	67.5	80.6	99.9	118.3	5.3	4.4
5	1	32	68.1	67.5	80.6	99.9	95.4	5.8	4.4
6	1	64	75.0	74.1	85.1	99.9	73.7	4.5	4.1
7	2	2	47.9	45.9	62.9	99.5	143.7	7.5	9.8
8	2	4	52.2	51.3	67.8	99.8	140.6	5.8	10.7
9	2	8	64.3	63.6	77.8	99.9	132.8	5.0	4.4
10	2	16	71.3	70.2	82.5	99.9	119.3	5.2	4.2
11	2	32	77.1	76.0	86.3	99.9	97.3	4.6	4.0
12	2	64	72.8	72.3	84.0	100.0	71.0	4.1	3.8
13	3	2	17.5	17.4	29.6	99.9	144.3	126.1	15.8
14	3	4	70.0	69.4	81.9	99.9	137.0	5.5	4.8
15	3	8	75.4	75.0	85.7	100.0	129.3	5.1	4.5
16	3	16	79.7	79.1	88.3	100.0	113.0	4.2	4.1
17	3	32	84.5	83.4	91.0	99.9	93.6	5.6	4.4
18	3	64	80.4	79.9	88.8	100.0	67.2	4.2	3.8

**Fig. 9** Pairs of ROI1 containing the acquired image, on the left, and the semantic segmentation prediction, on the right. The black line over the prediction image indicates the position of the strip estimated by the system. Pair 2*b* is blurred due to the steam content. Remaining water from the descaler process can be perceived in pairs 1*a*, 1*b*, and 2*a*. The reflections of the strip incandescence over the side guide structures are present in pairs 1*a* and 1*b*

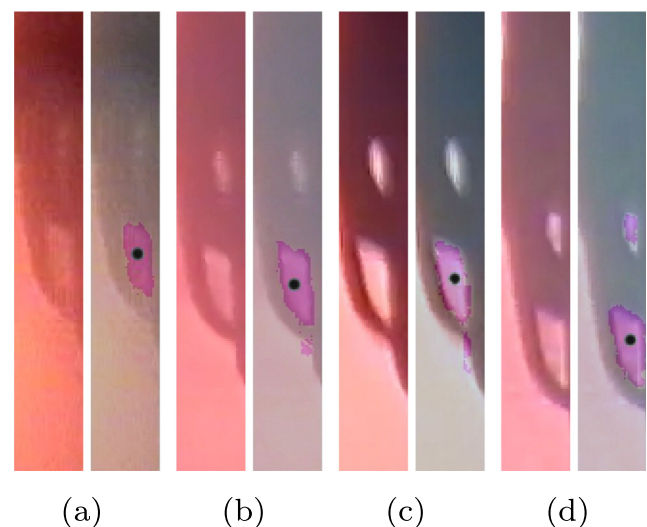


imbalance between the number of pixels in the reference polygon and the background. Secondly and most importantly, the tonality of the reference pixels is not distinct from other portions belonging to the background. These assumptions are confirmed by analyzing the results. The system exhibits a negligible false positive rate and a high false negative rate, which translates to mediocre recall and IoU while keeping a high specificity. In other words, the network tends to classify background pixels as a region of interest wrongly, but it does not misclassify background regions. Under these circumstances, the reference polygon could still be successfully identified after the morphological operations and grouping in connected components. Therefore, the reference position, which is derived from the polygon centroid, could be estimated.

Regarding the frame rate, since the input size is relatively small, all models are lightweight with execution rate above the requirements of the system. The model selection can be based solely in terms of average error and its dispersion. The most suitable network was Model 12, with 2 encoder/decoder pairs and 64 filters in each layer, attaining errors of only  $4.1 (\pm 3.8)$  mm and frame rate of 71.0 fps.

The results of the position estimation of the strip, by Model 3, and centroid, by Model 12, are also exemplified by the samples presented in Fig. 9, for ROI1, and Fig. 10, for ROI2. Both figures contain four pairs, from which we can observe the input of the region of interest on the left and the output on the right. The output image consists of the input image overlaid by the segmentation classification and the position estimated by the whole algorithm, including the morphological operations mentioned in Section 3.4.

The position estimated is shown as a black line, for the strip position, and a black dot, for the centroid position. The system demonstrated robustness by correctly placing the black line and black dot even when the segmentation step provides insufficient results. The proposed method can estimate the strip and centroid positions even in a complex environment, such as the steam presence (Fig. 9 pair 2*b*, and Fig. 10*a*, *b*, and *d*), mill structures reflecting the strip



**Fig. 10** Pairs of ROI2 containing the acquired image, on the left, and the semantic segmentation prediction, on the right. The black dot over the prediction image indicates the position of the centroid estimated by the system. Steam presence can be observed in pairs 1*a*, 1*b*, and 1*c*

incandescence (Fig. 9 pairs 1a and 1b) and water presence over the strip (Fig. 9 pairs 1a, 1b, and 2a).

The visualization of the expected and estimated values in each frame, as well as the absolute difference between them, are shown in Fig. 11. The top graph shows the results for the steel strip, while the bottom one refers to the centroid of the polygon. For more clarity, the images also present a magnified view of a signal stretch. As it can be observed, the estimations follow the expected values closely.

#### 4.2 Steel strip misalignment evaluation

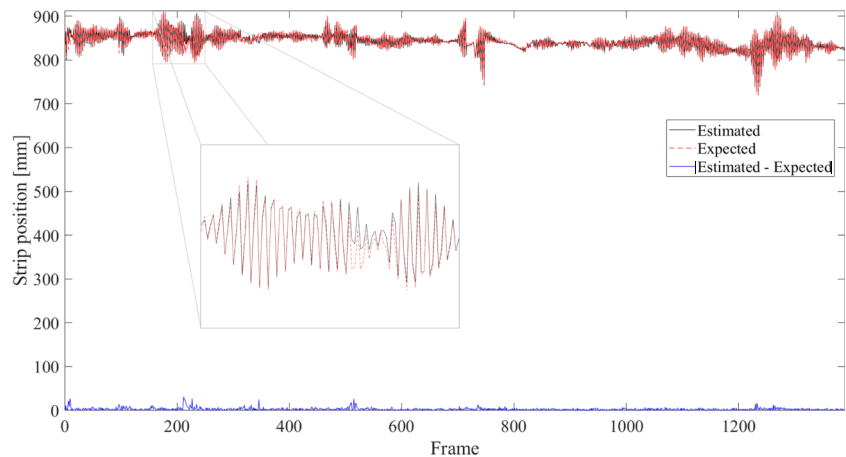
The goal of the present work is to determine the misalignment of the hot strip during the rolling process. This misalignment is assessed by the comparison between the steel strip location after the mechanical vibration influences are mitigated, and setpoint images. Graphical results of the system can be observed in Fig. 12. In the first image, Fig. 12a, it is shown the real deviance of the strip from its desired position and the estimated one by the developed system. The absolute difference between these quantities

is depicted in Fig. 12b, whereas a histogram portrays its distribution in Fig. 12c.

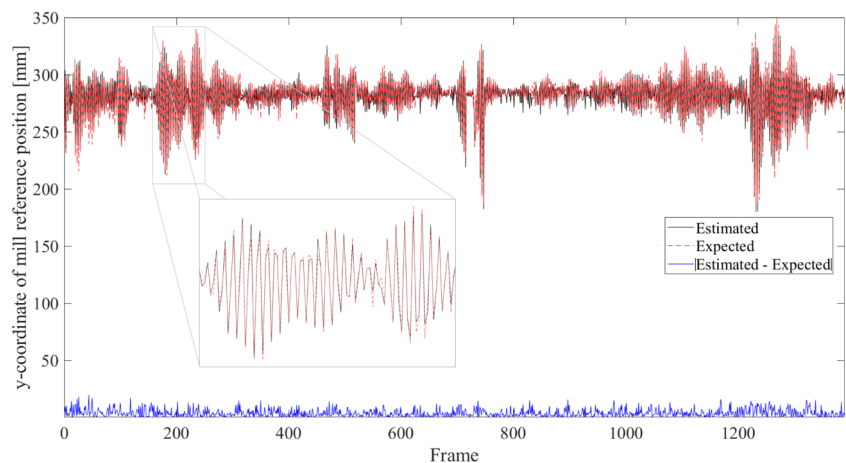
As can be observed in the histogram, the vast majority of the data deviates from the desired value in less than 5 mm. In fact, this deviation corresponds to merely 2 pixels and 97% of the samples lies within this range. The mean absolute error and standard deviation achieved by the method were  $2.06(\pm 1.7)$  mm. Moreover, the system is capable of performing the strip misalignment estimation in real-time, at a rate of 26.4 fps.

Although the manual operator performance cannot be accessed due to lack of data, the available literature presents studies of the human capability regarding reaction time. Studies of visuomotor reaction time (VMRT), in which participants executed a visuomotor reaction task in response to visual motion stimuli, showed that badminton players, table tennis athletes, and non-athletes presented VMRT of 244.2 ms, 258.4 ms, and 273.6 ms, respectively [16–18]. A rather distant approach for comparison, if a human operator had a similar performance of a badminton player, which corresponds to a reaction frequency of

**Fig. 11** Graphical result of position estimation in contrast to the expected values for **a** steel strip and **b** mill structure components

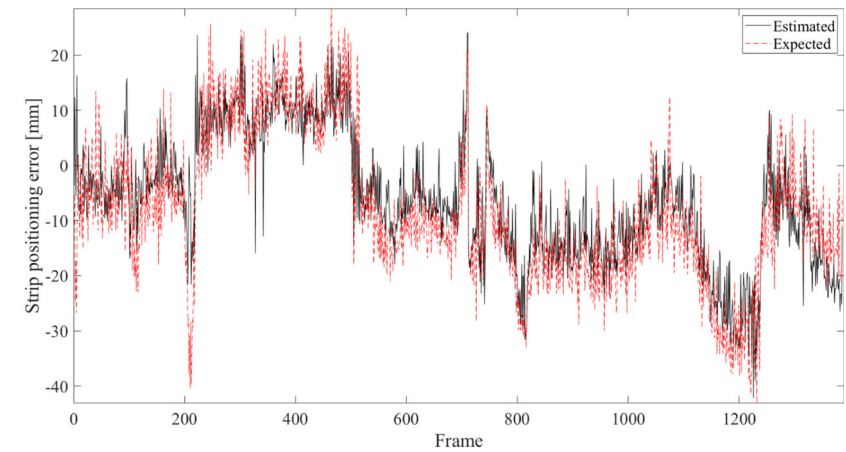


(a)

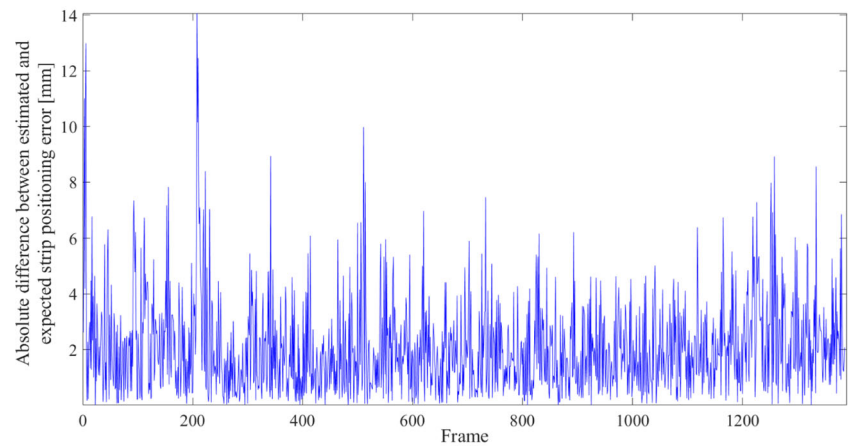


(b)

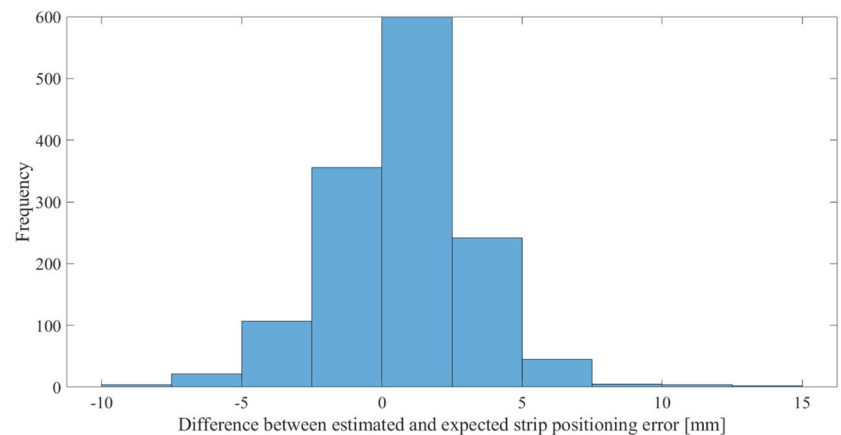
**Fig. 12** Results regarding strip position misalignment. **a** Estimated strip position error ( $Y_{error}$ ) in relation to the expected values. **b** Absolute difference between estimated and expected values. **c** Histogram of the differences between estimated and expected values



(a)



(b)



(c)

4.460 fps, the system's performance (26.4 fps) would overcome the human operator. Additionally, the majority of the data diverges from the expected value in less than 5 mm, which is equivalent to less than 2 pixels in the images. Considering that the strips are rolled at 10 m/s, it is reasonable to assume that the system also

outperforms the current manual operations in terms of precision.

Therefore, the presented approach can successfully estimate the steel strip misalignment with precision and response time well beyond the current human operator capabilities and, consequently, process specification.

## 5 Conclusions

The system presented in this paper aims to estimate steel strips positioning error. The method employs semantic segmentation to estimate the strip location from images of the process and to identify mill structure fixed parts used to attenuate the influences of camera vibration on the results. The identification of the strip by the convolution network achieved high accuracy for all tested models, in contrast to the mill structure, in which the identification by the CNN accomplished considerable results only for a few models. However, the application of morphological operations refined semantic segmentation predictions. As a result, the elected models achieved the strip location and mill elements parts with mean absolute errors of 3.4 ( $\pm$  4.0) mm and 4.1 ( $\pm$  3.8) mm, respectively.

Additionally, the method can successfully estimate the steel strip misalignment, presenting 97% of the estimated values within a 2-pixel margin. Also, concerning the positioning error, the mean absolute error attained by the system is 2.06 ( $\pm$  1.7) mm. Regarding the execution time, the method presented reduced computational cost per frame, with an approximate frame rate of 26 fps. All thoughts considered, the approach also proved to be a robust real-time solution as the dataset is composed of low-resolution images acquired in a complex environment. Future work will be carried out for the integration of the developed solution in a feedback control system designed to reject strip positioning errors.

**Acknowledgments** The research reported in this paper and carried out at the Budapest University of Technology and Economics was supported by the “TKP2020, National Challenges Program” of the National Research Development and Innovation Office (BME NC TKP2020).

**Funding Information** Open access funding provided by Budapest University of Technology and Economics.

## Compliance with ethical standards

**Conflict of interests** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Steckel mills – creative solutions for the metal industry (2000)
- Almotairi S, Kareem G, Aouf M, Almotairi B, Salem MAM (2020) Liver tumor segmentation in ct scans using modified segnet. *Sensors* 20(5):1516
- Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Patt Anal Mach Intel* 39(12):2481–2495
- Carruthers-Watt BN, Xue Y, Morris AJ (2010) A vision based system for strip tracking measurement in the finishing train of a hot strip mill. In: 2010 IEEE International conference on mechatronics and automation, IEEE, pp 1115–1120
- Chen FC, Jahanshahi MR (2017) Nb-cnn: deep learning-based crack detection using convolutional neural network and naïve bayes data fusion. *IEEE Trans Ind Electron* 65(5):4392–4400
- Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B (2016) The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3213–3223
- DeCost BL, Lei B, Francis T, Holm EA (2019) High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel. *Microsc Microanal* 25(1):21–29
- Ess A, Müller T, Grabner H, Van Gool LJ (2009) Segmentation-based urban traffic scene understanding. In: *BMVC*, vol 1, Citeseer, pp 2
- de Faria Lemos A, da Silva LAR, Furtado EC, de Paula H (2017) Positioning error estimation of steel strips in steckel rolling process using digital image processing. In: 2017 IEEE Industry applications society annual meeting, IEEE, pp 1–8
- Ferguson M, Ak R, Lee YTT, Law KH (2017) Automatic localization of casting defects with convolutional neural networks. In: 2017 IEEE International conference on big data (big data), IEEE, pp 1726–1735
- Ferreira ABS (2005) Adaptive fuzzy logic steering controller for a steckel mill. Ph.D. thesis, University of Johannesburg
- Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on computer vision and pattern recognition, IEEE, pp 3354–3361
- Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 580–587
- Hoang DT, Kang HJ (2019) Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cogn Syst Res* 53:42–50
- Hong S, Oh J, Lee H, Han B (2016) Learning transferrable knowledge for semantic segmentation with deep convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3204–3212
- Hülsdünker T, Ostermann M, Mierau A (2019) The speed of neural visual motion perception and processing determines the visuomotor reaction time of young elite table tennis athletes. *Frontiers in behavioral neuroscience*, pp 13
- Hülsdünker T, Strüder HK, Mierau A (2017) Visual motion processing subserves faster visuomotor reaction in badminton players. *Medicine and Science in Sports and Exercise* 49(6):1097–1110
- Hülsdünker T, Strüder HK, Mierau A (2018) Visual but not motor processes predict simple visuomotor reaction time of badminton players. *European Journal of Sport Science* 18(2):190–200

19. Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufier M, Verstockt S, Van de Walle R, Van Hoecke S (2016) Convolutional neural network based fault detection for rotating machinery. *J Sound Vib* 377:331–345
20. Konovalov YV, Khokhlov A (2013) Benefits of steckel mills in rolling. *Steel in Translation* 43(4):206–211
21. Kwon W, Kim S, Won S (2015) Active disturbance rejection control for strip steering control in hot strip finishing mill. *IFAC-PapersOnLine* 48(17):42–47
22. Lee SJ, Yun JP, Koo G, Kim SW (2017) End-to-end recognition of slab identification numbers using a deep convolutional neural network. *Knowl-Based Syst* 132:1–10
23. Lin G, Milan A, Shen C, Reid I (2017) Refinenet: multi-path refinement networks for high-resolution semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1925–1934
24. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3431–3440
25. Masci J, Meier U, Ciresan D, Schmidhuber J, Fricout G (2012) Steel defect classification with max-pooling convolutional neural networks. In: *The 2012 international joint conference on neural networks (IJCNN)*, IEEE, pp 1–6
26. Pham DL, Xu C, Prince JL (2000) Current methods in medical image segmentation. *Annual Review of Biomedical Engineering* 2(1):315–337
27. Rashed EA, Gomez-Tames J, Hirata A (2020) End-to-end semantic segmentation of personalized deep brain structures for non-invasive brain stimulation. *Neural Networks*
28. Roberts G, Haile SY, Sainju R, Edwards DJ, Hutchinson B, Zhu Y (2019) Deep learning for semantic segmentation of defects in advanced stem images of steels. *Scientific Reports* 9(1):1–12
29. Sadoughi M, Hu C (2019) Physics-based convolutional neural network for fault diagnosis of rolling element bearings. *IEEE Sensors J* 19(11):4181–4192
30. Sevak JS, Kapadia AD, Chavda JB, Shah A, Rahevar M (2017) Survey on semantic image segmentation techniques. In: *2017 International conference on intelligent sustainable systems (ICISS)*, IEEE, pp 306–313
31. Soukup D, Huber-Mörk R (2014) Convolutional neural networks for steel surface defect detection from photometric stereo images. In: *International symposium on visual computing*, Springer, pp 668–677
32. Trembl M, Arjona-Medina J, Unterthiner T, Durgesh R, Friedmann F, Schuberth P, Mayr A, Heusel M, Hofmarcher M, Widrich M et al (2016) Speeding up semantic segmentation for autonomous driving. In: *MLITS, NIPS Workshop*, vol 2, pp 7
33. Wang Z, Fan J, Jing F, Liu Z, Tan M (2019) A pose estimation system based on deep neural network and icp registration for robotic spray painting application. *The International Journal of Advanced Manufacturing Technology* 104(1-4):285–299
34. Wang ZH, Gong DY, Li X, Li GT, Zhang DH (2017) Prediction of bending force in the hot strip rolling process using artificial neural network and genetic algorithm (ann-ga). *The International Journal of Advanced Manufacturing Technology* 93(9-12):3325–3338
35. Wei Y, Chang-Qing S, Xiao-Jie G, Zhong-Kui Z (2017) Bearing fault diagnosis using convolution neural network and support vector regression. *DEStech Transactions on Engineering and Technology Research*
36. Xiao L, Lu M, Huang H (2020) Detection of powder bed defects in selective laser sintering using convolutional neural network. *The International Journal of Advanced Manufacturing Technology*, pp 1–12
37. Xie C, Wang J, Zhang Z, Zhou Y, Xie L, Yuille A (2017) Adversarial examples for semantic segmentation and object detection. In: *Proceedings of the IEEE international conference on computer vision*, pp 1369–1378
38. Xu ZW, Liu XM, Zhang K (2019) Mechanical properties prediction for hot rolled alloy steel using convolutional neural network. *IEEE Access* 7:47068–47078
39. Yang SS, He YH, Wang ZL, Zhao WS (2008) A method of steel strip image segmentation based on local gray information. In: *2008 IEEE International conference on industrial technology*, IEEE, pp 1–4
40. Youkachen S, Ruchanurucks M, Phatrapomnant T, Kaneko H (2019) Defect segmentation of hot-rolled steel strip surface by using convolutional auto-encoder and conventional image processing. In: *2019 10Th international conference of information and communication technology for embedded systems (IC-ICTES)*, IEEE, pp 1–5
41. Zhang W, Li C, Peng G, Chen Y, Zhang Z (2018) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech Syst Signal Process* 100:439–453
42. Zhang Z, Wu C, Coleman S, Kerr D (2020) Dense-inception u-net for medical image segmentation. *Computer Methods and Programs in Biomedicine*, pp 105395

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.