Computer Vision as Key to an Automated Concrete Production Control

Max Coenen[1], Max Meyer[2], Dries Beyer[1], Christian Heipke[2], Michael Haist[1]

[1]Institute of Building Materials Science, Leibniz University Hannover, Germany
[2]Institute of Photogrammetry and GeoInformation, Leibniz University Hannover, Germany

[m.coenen, d.beyer, haist]@baustoff.uni-hannover.de, [meyer, heipke]@ipi.uni-hannover.de

Abstract -
The need to reduce CO_2 emissions from concrete leads to increasingly complex mix designs involving e.g. CO_2 reduced cements, recycled materials, and various chemical additives. This complexity results in a larger sensitivity of the concrete to unpredictable fluctuations in both, the base material properties and in boundary conditions such as temperature and humidity during the production process. Digital sensor systems and quality control schemes are considered as key to counteract this problem by enabling an automated production control. As contribution towards this goal, this paper investigates the research question whether Computer Vision can be used for the predictive characterisation of raw materials (here: of concrete aggregates) and of the fresh concrete quality during the mixing process. In particular, we propose the usage of imaging sensors for the observation of both, aggregate material and the flow behaviour of fresh concrete during the mixing process, and present deep learning methods for the prediction of granulometric and rheological properties from the image observations, respectively. Incorporating such systems into the concrete production process enables the facilitation of a digital control loop for ready-mixed concrete production by allowing an in-line reaction to raw material fluctuations and to deviations of the concrete from the target properties.

Keywords -
Concrete 4.0, Computer Vision, Deep Learning, image-based granulometry, image-based rheology, CNN, ViT

1 Introduction

Concrete, which is the most used (building) material worldwide, contributes approximately 8% to the global CO_2-emissions. The urgent need to reduce CO_2-emissions of the concrete production results in increasingly complex concrete mixtures: Historically, being a 3-component-material (cement, water, and aggregates), modern recipes contain a large variety of different components, including e.g. CO_2 reduced cements (with various sub-components), recycled aggregates, and various chemical additives (super-plasticiser, flow-retarder, etc.) [1]. However, the growing complexity of the concrete mixtures results in a strongly increased sensitivity of the concrete properties to unpredictable fluctuations in the raw materials and the production processes, leading to diminished robustness of the concrete, to difficulties in reaching the desired concrete quality and ultimately, to higher rejection rates. These challenges are extended by the fact that until
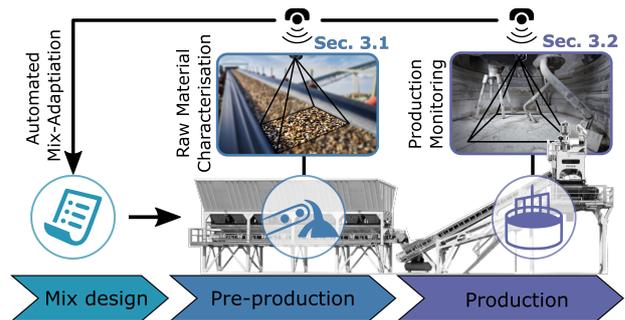

Figure 1. High-level overview on the proposed concrete production control scheme using computer vision based methods for the in-line characterisation of raw materials and fresh concrete quality.

today, the mixture design of the concrete and its production are mostly empirically driven, leading to additional difficulties due to the lack of experience with such new and complex concrete compositions. As a result, the extensive use of new, environmentally and resource-friendly building materials is hampered by the inability of controlling and guaranteeing the desired concrete properties during the production process.

In the authors' opinion, a digital mapping of the essential process steps within the concrete production chain is key for a transition towards a more sustainable high quality concrete construction industry [2]. While in many manufacturing industries, digitisation and automation have enabled a strong increase in productivity over the last decades, the developments in the construction industry stagnated during this period, resulting in the construction industry - and here, especially the concrete sector - to be one of the least digitised industries of the global economy [3]. In the context of the topic addressed by this paper, the lack of digitisation and automation is for instance reflected in the conventional, batch-based, and manual methods that are currently still used as standard for the characterisation of raw materials (e.g. manual sieving for estimating the particle size distribution of aggregates) and the quality control of the concrete production (e.g. slump flow or rheometer tests for determining fresh concrete properties). In order to overcome current limitations and to enable an

automated and improved concrete production, this paper presents computer vision based methods for an in-line monitoring of the raw materials used for concrete production as well as for an in-line quality control of the fresh concrete properties during the mixing process (cf. Fig.1). Given real-time information about the raw material characteristics before mixing (pre-production) allows to account for deviations from the expected properties by adapting the mix design accordingly, while real-time information on the fresh concrete properties (during production) allows for taking measures to assure that the target quality is reached. In particular, to investigate the suitability of imaging systems for the characterisation of building materials, we make the following contributions in this paper:

- We propose a Vision Transformer (ViT) based method for the visual granulometry of concrete aggregates from image observations.

- We present a Convolutional Neural Network (CNN) based approach for the prediction of fresh concrete properties from images acquired during the mixing process of the concrete and mix design information.

- In the experimental part of this paper, we demonstrate and evaluate the performance of the proposed methods on challenging real-world data sets.

2 Background

2.1 Raw material characterisation

The size distribution (formally known as grading curve) of the aggregates used for concrete production significantly influences many important concrete properties including e.g. the consistency, workability, and segregation tendency of the concrete in the fresh state as well as the compressive strength and the durability in the hardened state [4]. As a consequence, the size distribution of the aggregates has to be considered during mix design (typically by considering the aggregates' *k-value*), since it affects the water demand and the amount of required super-plasticizer of the concrete composition. In practice, however, the size distribution is usually determined on a low-frequency basis by mechanical sieving of small sample batches of aggregates (a few kilograms), which are considered as representative for large amounts (a few tons). As a consequence, deviations of the actual material used for concrete production from the test sample cannot be accounted for.

A camera based sensor setup, observing the material while being transported over the conveyor belt into the mixer, and algorithms that are able to predict the size distribution from the image observations allow to consider the actual aggregate grading curves in the mix design for the production of each concrete batch. In the literature, the derivation of size distributions from images of aggregate is often approached in a two-step procedure. In this context, the first step consists of a segmentation of individual particles, e.g. based on grayscale thresholding, edge detection, or watershed transformations in early approaches [5, 6], while modern procedures apply deep learning based methods for instance segmentation or panoptic segmentation of aggregate particles [7, 8]. In a second step, the particle size distribution is derived from the segmentation results. However, this procedure suffers from partial occlusions and an often insufficient spatial image resolution, and needs an explicit conversion from the two-dimensional segmentations to volumetric entities [5], introducing inaccuracies for the task of estimating size distributions.

In contrast to the described object-based procedure, statistical approaches avoid the explicit detection and modelling of individual instances by relying on global image statistics in order to predict the size distribution directly from the raw image. In this context, Olivier et al. [9] and Coenen et al. [10] propose to learn a CNN in order to distinguish different predefined particle size distributions. However, in this way, only a classification of a discrete set of grading curves is possible. In order to overcome this limitation, CNN based approaches for the prediction of the continuous percentiles defining the size distribution were presented in [11]. While the latter approaches used CNN-architectures, recently the application of transformer based models for vision tasks has shown great potential and encouraging results [12]. Since it was shown in the literature, that Vision Transformers (ViT) perform equal to and even outperform CNN-based methods for the task of grading curve prediction [13], we follow the recent success of transformer-based models and choose a ViT approach for the determination of concrete aggregate size distributions.

2.2 Fresh concrete characterisation

In current practice, the quality inspection of fresh concrete is mainly conducted offline, i.e. after the mixing and production process, using empirical test methods based on small batch samples taken from the concrete. One of these methods is, for example, the slump flow test [14]. This test involves applying a force to a fresh concrete sample and measuring how much it spreads. The manually measured slump flow diameter δ is an indicator of the consistency of the fresh concrete. However, the manual measurement entails a higher degree of inaccuracy due to possible human error. Furthermore, at this stage of the production process, only very limited control of the concrete properties remains possible. For this reason, an inline quality assessment during the mixing process is desirable, since it enables a real-time reaction on potential deviations from the target properties. Besides, caused by the hydration process of concrete, its properties can change significantly between the time of mixing and the time of the actual

placement of the concrete. Therefore, it is also striven for a prediction of the concrete properties at the time of placement by modelling the time-dependent behaviour of the concrete. In addition to the consistency, the fresh concrete can also be characterised by its rheological parameters, namely the plastic viscosity μ and the yield stress τ_0 of a Bingham model which is typically used to describe the non-Newtonian properties of fresh concrete. Founded on fluid dynamic laws, these properties determine the flow behaviour of the fluid under specific boundary conditions, e.g. during a mixing process. For the prediction of fresh concrete properties, two central procedures can be found in the literature. One approach is to make use of the information of the mix design on the basis of which machine learning methods are trained to predict the target parameters. In [15], the information from the mix design is additionally extended by temporal information to enable a time-dependent prediction. Although these approaches are promising, they heavily depend on the quality of the information from the mix design which, however, are often subject to larger inaccuracies, e.g. caused by unknown fluctuations in the raw material properties. The second approach is to use sensor observations of the fresh concrete, e.g. in the form of image observations, and to establish deep learning methods to predict the fresh concrete properties on the basis of the observed data. In this context, [16] used a camera setup to observe the concrete flow at the outlet of a mixing truck and a CNN-based method was proposed to learn a mapping between the sequential image data and the fresh concrete's rheological properties. However, compared to channel flow, a significantly more complex behaviour is expected for the dynamic flow of the concrete during mixing. In [17], a method for predicting the rheology of Bingham fluids from stereo-images and 3D surface reconstructions of the material during the mixing procedure was presented. In [18], a similar approach using a CNN and a long-short-term memory network has been proposed. Although these approaches are not prone to possible uncertainties in the mix design, they do not take into account the time dependency of the properties. In this paper, we extend the CNN-based method of [17] and propose a novel procedure for time-dependent fresh concrete characterisation during the mixing process based on stereo-camera observations, temporal information, and mix design. Investigations whether the deployment of ViT-based models instead of a CNN architecture can improve the performance are planned in future work.

3 Methodology

3.1 Image-based granulometry

Given a three-channel (RGB) image I depicting concrete aggregates, we aim at automatically deriving the grading curve $G = [p_1, p_2, ..., p_N]$. Per definition, the grading curve is a histogram in which grain size intervals (bins) are represented on the abscissa (x-axis) and the quantity proportion is shown on the ordinate (y-axis). Consequently, G is parameterised by a vector whose elements p_j correspond to the histogram percentiles of each grain size interval $j = [1...N]$. In this representation, each percentile is a continuous variable with $\{p_j \in \mathbb{R} \mid 0 \leq p_j \leq 1\}$ under the constraint $\sum p_j = 1$. As a result, mapping the image I to the grading curve G corresponds to a constrained multi-regression problem of the individual percentiles p_j, in which the determined percentiles must sum up to 1. In order to tackle this problem, we make use of a Vision Transformer (ViT) based architecture [12] acting as mapping function $f : I \rightarrow G$. As is shown in Fig. 2, the ViT takes a single image as input and decomposes it into a sequence of n non-overlapping image patches $x_i \in \mathbb{R}^{h \times w}$ which are transformed into 1D tokens $z_i \in \mathbb{R}^d$ of length d using a linear projection $\mathbf{E}$. The sequence of tokens $\mathbf{z}^0 \in \mathbb{R}^{(n+1) \times d}$ with

$$\mathbf{z}^0 = [z_{\text{cls}}, \mathbf{E}z_1, \mathbf{E}z_2, ..., \mathbf{E}z_n] + \mathbf{p} \qquad (1)$$

then serves as input to a transformer encoder architecture [19]. As indicated in Eq. 1, a learnable classification token z_{cls} is prepended to the sequence, whose representation at the final layer of the encoder is used as input embedding for the output layer.
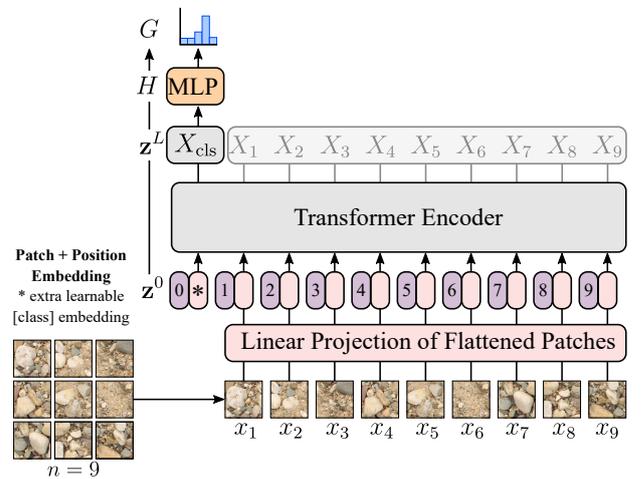

Figure 2. Structure of the ViT (adapted from [12])

Furthermore, a learnable position embedding $\mathbf{p} \in \mathbb{R}^{n \times d}$ is added to the tokens (cf. Eq. 1) in order to retain positional information throughout the permutation invariant self-attention operations of the encoder. The tokens are passed through the transformer encoder which consists of a stack of $l = 1...L$ residual layers, each comprising Multi-Head Self-Attention (MSA) [19], layer normalisation (LN), and Multi-Layer Perceptron (MLP) blocks. The

output of the last layer of the transformer encoder is denoted as embedding $\mathbf{z}^L = [X_{\text{cls}}, X_1, X_2, ..., X_n]$, where n is the total number of patch tokens and X_{cls} and $X_1, ..., X_n$ correspond to the embedded class token and patch tokens, respectively. Finally, a MLP head H is used on top of the transformer encoder and typically produces the prediction output based on the final encoded class token embedding X_{cls}. We make use of the softmax-function as final activation in H, returning $j = 1...N$ output values representing the size distribution G which consequently comply with the constraint $\sum p_j = 1$. For training, the loss for image I can be written as

$$L_{\text{cls}} = D\left(H(X_{\text{cls}}), Y_I\right), \qquad (2)$$

where $H(X_{\text{cls}})$ is the output of the final prediction head for the class embedding X_{cls}, Y_I is the reference for image I, and $D(\cdot, \cdot)$ is a distance function. In this work, we apply the Kullback-Leibler divergence D_{KL} as measure for $D(\cdot, \cdot)$ which computes the similarity between the predicted and the reference grading curves according to

$$D_{\text{KL}} = \sum_{j=1}^{N} p_j \cdot \log\left(\frac{p_j}{\hat{p}_j}\right), \qquad (3)$$

where p_j and $\hat{p}_j$ are the reference and the predicted percentiles of the size distribution, respectively.

3.2 Image-based fresh concrete characterisation

Until today, existing test methods for the determination of fresh concrete properties are exclusively applicable in post-production. As a result of the inability to estimate the fresh concrete properties already during the production process, an adaptation and adjustment of the concrete in case of deviations from the desired properties is not possible in current practice. To overcome this limitation, in the paper we present an image based method for the inline prediction of fresh concrete quality during the production, i.e. during the mixing process in this paper. In addition to the flow behaviour, which contains valuable information about the rheological properties, previous studies have shown that the 3D surface carries ancillary information able to support the task of rheological characterisation [17]. For this reason, we propose a stereo-camera setup observing the flow behaviour of the concrete during mixing, allowing to reconstruct the 3D surface of the concrete via dense stereo-matching and triangulation. More specifically, we compute a digital elevation model (DEM) D together with an orthophoto O for each stereo image pair. Both entities D and O are used as input to a Convolutional Neural Network (CNN) with the goal to predict a state vector $C = [\delta, \tau_0, \mu]$ representing the fresh concrete characteristics, namely the slump flow diameter δ (defining the concrete consistency), and the Bingham parameters

τ_0 and μ, describing the rheological properties. As additional input information, we introduce the composition data of the mix design m, consisting of information about the water-cement ratio, the grading curve, the paste content, the additive content and the time difference between the start of the mixing process and the image acquisition, and append it to the latent feature embedding produced by the CNN in a late-fusion manner. A high-level overview on the procedure is shown in Fig. 3. Since deeper networks tend to overfit, the CNN consists of only 7 convolutional layers. Each layer has a kernel size of 5x5 and a stride of 2, followed by a batch normalisation and a Rectified Linear Unit (ReLU) activation function. The convolutional layers are followed by three fully connected (FC) layers, each with a leaky ReLU activation function using a slope of 0.2. O and D are passed through the convolutional layers, producing the flattened feature embedding $z_{O,D}$. By concatenating $z_{O,D}$, the mixture design information m, and the time epoch Δ_t we obtain the feature vector F, which is used as input for the FC layers. Here, Δ_t represents the temporal difference between the acquisition time of the input image pair and the point in time where the consistency is to be determined. As a consequence, the target state vector C becomes time-dependent, and therefore, the FC-layers finally map the feature vector to the three time-dependent output parameters δ_{Δ_t}, τ_{0,Δ_t} and μ_{Δ_t}.
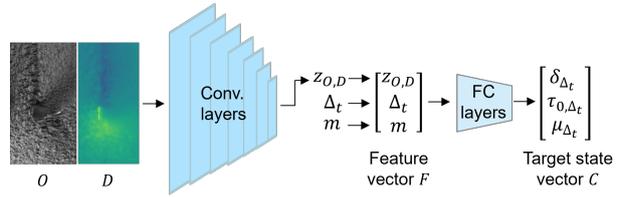

Figure 3. Architectural structure of the CNN.

For optimising the network weights ω the Mean Square Error (MSE) is used as loss function, which is iteratively minimised during training. In this context, the loss is computed for a mini-batch consisting of N samples, each associated with the state vector C containing the $k = 1...K$ target parameters y^k, where $K = 3$ and $N = 32$ in this paper. For loss calculation, the squared differences between the reference values y^k and the predicted values $\hat{y}^k$ are determined and averaged over all parameters and samples in a batch, such that

$$L_{\text{MSE}}(\omega) = \frac{1}{N \cdot K} \sum_{k=1}^{K} \sum_{n=1}^{N} (y_n^k - \hat{y}_n^k)^2. \qquad (4)$$

To prevent over-fitting, weight decay is used during training. This method penalises large weights by a factor λ (in this paper $\lambda = 0.001$) and leads to an additional term in

the final loss function

$$L(\omega) = L_{\text{MSE}}(\omega) + \lambda \cdot \sum \omega^2. \qquad (5)$$

4 Experiments

4.1 Image-based granulometry

Test setting: For the experimental evaluation of the proposed method for particle size distribution estimation, we make use of the publicly available *Deep Granulometry data set*[1]. The data set consists of images showing concrete aggregate particles and reference data of the particle size distribution (grading curves) associated to each image. The data consists of two independent sets of images, the *Coarse Aggregate Data* (D_{coarse}) and the *Fine Aggregate Data* (D_{fine}). Both contain approximately 1700 images associated to one of 34 different particle size distributions. While the data in D_{coarse} shows aggregates with particle sizes ranging from 0.1 to 32 mm and provides reference percentiles for $N = 9$ bins with upper bounds of 0.25, 0.5, 1, 2, 4, 8, 16, 32.5, 63 [mm], the D_{fine} data contain fine material with grain sizes between 0 and 2 mm and with references for $N = 6$ bins, namely 0.063, 0.125, 0.25, 0.5, 1.0, 2.0 [mm]. We train networks for each set of images (fine and coarse) individually. Due to computational reasons we do not make use of the full image resolution in which the data is provided but we downsample the images to a size of 512x704 [px] for the D_{coarse} data, corresponding to a ground sampling distance (GSD) of 0.5 mm, and to a size of 480x480 [px] for the D_{fine} data, corresponding to a GSD of 0.1 mm. We adapt the *hybrid ViT-Base* architecture according to the definition in [12], i.e. we feed the image to a convolutional module and form the input sequence for the transformer based on the resulting feature maps (cf. [13] for details). For the architecture of the transformer encoder, we chose $L = 12$ layers, each comprising 12 multi-head-self-attention (MSA) modules. For evaluation, we split the data sets into two subsets and perform a two-fold crossvalidation. In order to reduce overfitting effects, we make use of random geometric and radiometric image augmentations during training.

Results: In order to assess the performance of the predicted particle size distributions, we compute the mean absolute error ε_p over the individual percentiles p_j of the different bins. Furthermore, for testing the grading curve predictions as a whole, we compute the average ε_H of the discrete *Hellinger distances* $D_H \in [0, 1]$, used as measure of the similarity between the reference grading curves G

and the predicted grading curves $\hat{G}$, with

$$D_H(G, \hat{G}) = \frac{1}{\sqrt{2}} \sqrt{\sum_j (\sqrt{p_j} - \sqrt{\hat{p}_j})^2}. \qquad (6)$$

Here, values close to 0 represent highly similar distributions and values close to 1 highly different distributions. For concrete mix design, often the k-value of the aggregates, which is computed as the sum of the percentiles of the cumulative grading curve representation, and which represents a measure for the granularity of the aggregate, is used to adjust the water content of the composition. We therefore also compute the k-value from the predicted size distributions, compare it to the reference values, and report the mean absolute error $\varepsilon_{\text{k-value}}$. Tab. 1 contains the results for the described metrics on both data sets. As is visible from the table, the percentile-wise predictions lead to an average error of $\varepsilon_p = 4.21\%$ on the fine aggregate data, and to only 1.57% on the coarse aggregate data. As a consequence, also the values for the average Hellinger distance and the average errors of the k-values are comparably smaller on the D_{coarse} data.

In order to gain more detailed insights into the distribution of the respective errors, Fig. 4 shows the cumulative histogram of the absolute percentile errors (Fig. 4a) as well as a plot of the reference k-values vs. predicted k-values on the fine (Fig. 4b) and the coarse (Fig. 4c) data sets. Regarding the percentile-wise prediction errors it can be seen from Fig. 4a, that approx. 90% of all values of the fine aggregate data are predicted with an error of less than 10%, and nearly 100% of all values on the coarse aggregate data. W.r.t. the resulting k-values, the most important indicator for the concrete production, it can be seen that the predicted values show some fluctuations but, on average, are very close to the reference values, demonstrating a highly promising suitability of the proposed method for a future implementation in practice.

Table 1. Quantitative results of the image based granulometry on the two data sets D_{coarse} and D_{fine}.

	ε_p [%]	ε_H	$\varepsilon_{\text{k-value}}$
D_{fine}	4.21	0.127	0.187
D_{coarse}	1.57	0.071	0.145

4.2 Image-based fresh concrete characterisation

Test setting: For training and testing of the method proposed for the determination of fresh concrete properties, we acquired an extensive data set using a similar setup as Ponick et al. [17]. More specifically, we designed a *surrogate-mixing system* consisting of a linear

---

[1]https://doi.org/10.25835/61y9peiq

(a) Cumulative histogram of the absolute percentile-wise prediction errors.

(b) Prediction results for the k-values on the D_{fine} data.

(c) Prediction results for the k-values on the D_{coarse} data.
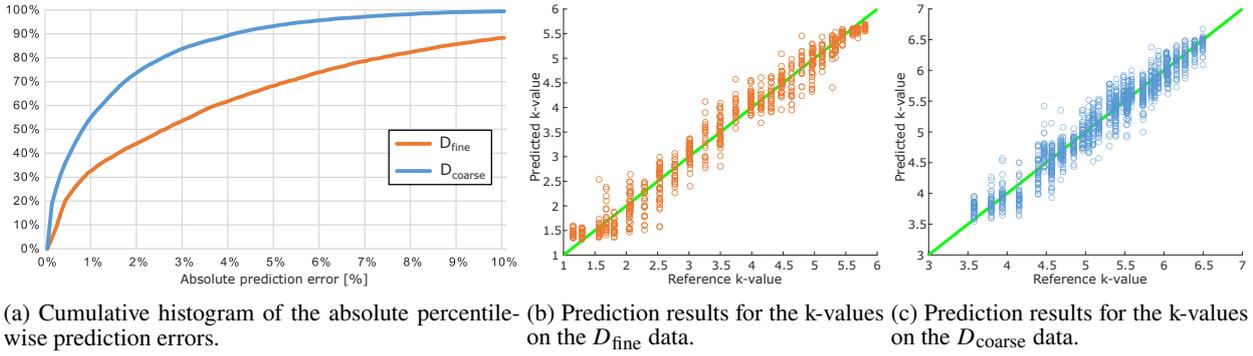
Figure 4. Performance results for the predictions of the aggregate particle size distributions.

mixing geometry in form of a channel, which is filled with fresh concrete and in which a single mixing paddle moves back and forth along a linear trajectory in order to simulate a simplified mixing process of a concrete mixer. During movement of the paddle, a stereo camera setup (2xGrasshopper 3 USB cameras with a focal length of 8 mm) is used to observe the motion behaviour of the concrete by acquiring images of size 1920x1200 [px]. Fig. 5 shows the schematic test setup.
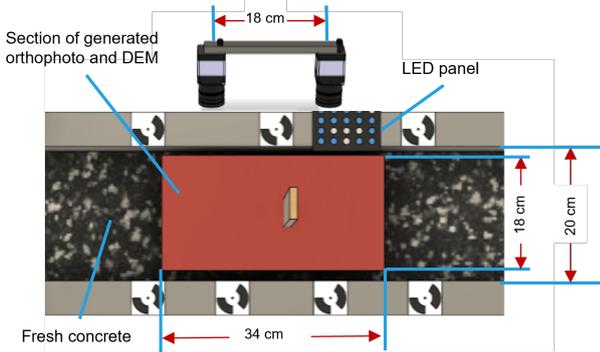

Figure 5. Top view on the *surrogate mixing system* used for generating the fresh concrete mixing data.

For the experiments in this paper, a total of 45 concretes with different mix compositions were selected. In particular, we generated different mixture designs by varying the parameters of water-cement ratio, paste content, grading curve, cement type, sand-lime powder content and additive content. In order to generate reference values for the resulting fresh concrete properties, independent measurements were conducted in coordination with the image data acquisition of the mixing behaviour. The slump flow diameter δ was determined from the slump flow test [14], yield stress τ_0 and viscosity μ were determined using an *eBT-V* rheometer from *Schleibinger*. The first slump flow test for each concrete was made directly after finishing the mixing process. The slump test and the rheometer mea-

surement are independent of each other and were repeated several times at intervals of about 30 minutes in order to account for the time-dependent properties of the concrete.

Regarding image acquisition, 14 runs were recorded for each concrete, respectively, where one run corresponds to six back-and-forth movements of the mixing paddle. In run 1-7, the paddle moved with a velocity of 200 $\frac{mm}{s}$ and images were recorded with 30 frames per second (fps). In run 8-14, the velocity was set to 450 $\frac{mm}{s}$ while the frame rate was increased to 60 fps. To account for effects of different mixing velocities and frame rates, we include both information in m. A number of 1300 image pairs were captured during each run. For the generation of O and D, only the image pairs in which the paddle was visible in both images were used, resulting in a total amount of approx. 314k images in the data set. Note, that due to the dynamic nature of the scene, the images must be captured strictly at the same time for a valid generation of O and D from the stereo-pairs. To verify the synchronisation of the cameras, a LED panel with 20 LEDs was used, allowing to generate time stamps for each image at millisecond intervals. By assuming that the concrete properties remain constant during the relatively short time period of one run (approx. 44 and 22 sec), we associate each image pair within one run with the time stamp of the central image pair of the run. Consequently, Δ_t represents the time difference between the associated time stamp of the image pair and the point in time of the respective reference measurement.

Training: Since the three reference values for the multi-task output of the CNN are generated by two independent measurements, we created multiple reference combinations for each set of inputs O and D during training. For this purpose, we computed all possible reference combinations for each concrete and randomly assigned all entities O and D of a concrete to one of these combinations. As a consequence, Δ_t consists of two values, one for δ and one for τ_0 and μ, respectively. Training was performed by applying a five-fold cross-validation. The 45 concretes were divided into 5 sets of 9

concretes each. First, we sorted the concretes according to the length of the first slump flow diameter δ_1 and formed 3 groups (with the 15 longest δ_1, the 15 central δ_1 and the 15 smallest δ_1, respectively). Subsequently, three concretes of each group were randomly assigned to one of the 5 sets to ensure a balanced distribution of the concretes. In each cross-validation step, one of the sets was used as the test set. The validation set, consisting of 5 concretes, was formed randomly from the concretes of the other sets, again by taking δ_1 into account. The remaining 31 concretes form the training set.

Results: After training, the network is evaluated on the test sets. In order to assess the performance of the network, the absolute and relative error of the predicted parameters are determined for each pair of O and D in the test set. Subsequently, we compute the mean absolute deviation ε_{abs} and the mean relative deviation ε_{rel} for each parameter. The results are shown in Tab. 2. The results improve if the predictions of the same label combination (and different inputs) in a run are averaged beforehand and then ε_{abs} and ε_{rel} are calculated (on average approx. 41 predictions were averaged). These results are shown in brackets. With 6.1% of relative error for the prediction of the slump flow diameter, the results are already in a very promising range, especially considering that the slump flow test, which is used as reference method, is a manual measure associated with a considerable uncertainty. Compared to the slump flow, the results obtained for the yield stress and the viscosity are distinctly less precise. The reason for this is probably that the rheometer measurements are actually intended for a testing directly after the mixing process, as the concrete is more fluid at this point. With increasing time and increasing viscosity, the uncertainty of the rheometer measurements increases.

Table 2. Performance metrics of the CNN for the prediction of fresh concrete properties.

	ε_{rel}	ε_{abs}
δ_{Δ_t}	6.4% (6.1%)	2.8 cm (2.7 cm)
τ_{0,Δ_t}	28.3% (26.9%)	53.6 Pa (50.6 Pa)
μ_{Δ_t}	43.8% (42.4%)	18.5 Pa·s (17.8 Pa·s)

By using the time difference as input, the CNN also implicitly learns a time-dependent prediction model for the fresh concrete properties. This enables the prediction of the properties of the concrete at the time of placement during the mixing process (provided the time between mixing and placement is known). Fig. 6 shows examples of the temporal predictions for the slump flow diameter. On the y-axis the consistency parameters are plotted and

the x-axis represents the time difference Δ_t between the time of image acquisition and the time at which the target parameters are to be determined. For the reference values, the precision 2.46 cm of the slump test is given as error bar [14]. Since some reference measurements were taken before the image acquisition, the x-axis in the figures starts with a negative value. The results show that the network is able to learn the time-dependent behaviour of the concrete consistency represented by the slump flow diameter.
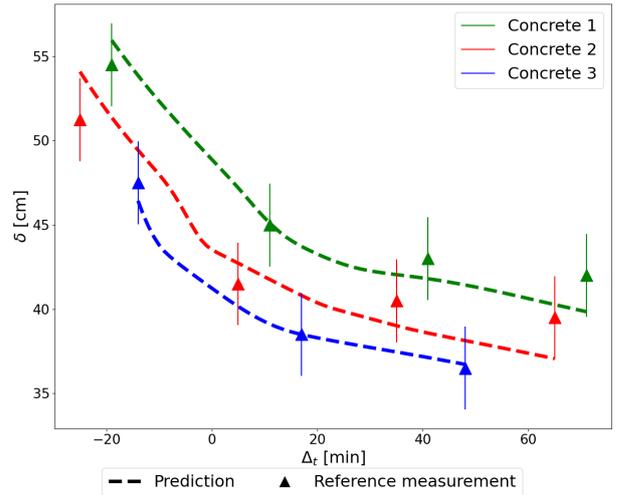

Figure 6. Three examples for the prediction of the time-dependent behaviour of the slump flow δ.

5   Conclusion

In this paper, we presented computer vision based methods for the characterisation of concrete aggregate and fresh concrete from image sequences acquired during the mixing process. The results obtained on challenging data sets demonstrate a promising performance for both, predicting the precise particle size distribution of fine and coarse aggregate particles, and deriving relevant fresh concrete properties. In the future, we aim at applying both methods under real-world conditions within a concrete production facility in order to integrate these methods into a control loop for the concrete production. In particular, strategies for an in-line adaptation of the mix design and countermeasures for a control of the fresh concrete properties during mixing will be developed on the basis of the approaches proposed in this paper.

Acknowledgements

References

[1] I. González-Taboada, B. González-Fonteboa, F. Martínez-Abella, and N. Roussel. Robustness of self-compacting recycled Concrete: Analysis of Sensitivity Parameters. *Materials and Structures*, 51(8), 2018. doi:10.1617/s11527-017-1136-1.

[2] M. Haist, T. Schack, M. Coenen, C. Vogel, D. Beyer, and C. Heipke. Concrete 4.0 - Sustainable Concrete Construction with Digital Quality Control. In *ce/papers 6(6)*, pages 1555–1562, 2023. doi:10.1002/cepa.2965.

[3] B. Green. *Productivity in Construction: Creating a Framework for the Industry to Thrive*. Chartered Institute of Building (CIOB), 2016.

[4] W. Fuller and S. Thomson. The Laws of Proportioning Concrete. *Transactions of the American Society of Civil Engineers*, 59(2):67–143, 1907.

[5] E. Hamzeloo, M. Massinaei, and N. Mehrshad. Estimation of Particle Size Distribution on an Industrial Conveyor Belt using Image Analysis and Neural Networks. *Powder Technology*, 261:185–190, 2014. doi:10.1016/j.powtec.2014.04.038.

[6] F. C. Lira and P. Pina. Grain Size Measurement in Images of Sands. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 371–374, 2006.

[7] A. Soloy, I. Turki, M. Fournier, S. Costa, B. Peuziat, and N. Lecoq. A Deep Learning-Based Method for Quantifying and Mapping the Grain Size on Pebble Beaches. *Remote Sensing*, 12(21), 2020. doi:10.3390/rs12213659.

[8] M. Coenen, T. Schack, D. Beyer, C. Heipke, and M. Haist. ConsInstancy: Learning Instance Representations for Semi-Supervised Panoptic Segmentation of Concrete Aggregate Particles. *Machine Vision and Applications*, 33(57), 2022. doi:10.1007/s00138-022-01313-x.

[9] L. E. Olivier, M. G. Maritz, and I. K. Craig. Deep Convolutional Neural Network for Mill Feed Size Characterization. *IFAC-PapersOnLine*, 52(14):105–110, 2019. doi:10.1016/j.ifacol.2019.09.172.

[10] M. Coenen, D. Beyer, C. Heipke, and M Haist. Learning to Sieve: Prediction of Grading Curves from Images of Concrete Aggregate. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-2-2022*, pages 227–235, 2022. doi:10.5194/isprs-annals-V-2-2022-227-2022.

[11] N. Lang, A. Irniger, A. Rozniak, R. Hunziker, J. D. Wegner, and K. Schindler. GRAINet: Mapping Grain Size Distributions in River Beds from UAV Images with Convolutional Neural Networks. *Hydrology and Earth System Sciences*, 25(5):2567–2597, 2021. doi:10.5194/hess-25-2567-2021.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, and et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations (ICLR)*, 2021.

[13] M. Coenen, D. Beyer, and M. Haist. Granulometry Transformer: Image-based Granulometry of Concrete Aggregate for an Automated Concrete Production Control. In *Proceedings of the 2023 European Conference on Computing in Construction (EC3)*, 2023. doi:10.35490/EC3.2023.223.

[14] EN 12350-5. Testing Fresh Concrete - Part 5: Flow Table Test. European Committee for Standardization, 2019.

[15] I. Navarrete, I. La Fé-Perdomo, J. A. Ramos-Grez, and M. Lopez. Predicting the Evolution of Static Yield Stress with Time of blended Cement Paste through a Machine Learning Approach. *Construction and Building Materials*, 371, 2023. doi:10.1016/j.conbuildmat.2023.130632.

[16] M. Coenen, C. Vogel, T. Schach, and M. Haist. Deep Concrete Flow: Deep Learning based Characterisation of fresh Concrete Properties from Open-Channel Flow using Spatio-Temporal Flow Fields. *Construction and Building Materials*, 411, 2024. doi:10.1016/j.conbuildmat.2023.134809.

[17] A. Ponick, A. Langer, D. Beyer, M. Coenen, M. Haist, and C. Heipke. Image-Based Deep Learning for Rheology Determination of Bingham Fluids. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2022*, pages 711–720, 2022. doi:10.5194/isprs-archives-XLIII-B2-2022-711-2022.

[18] L. Yang, X. An, and S. Du. Estimating Workability of Concrete with Different Strength Grades based on Deep Learning. *Measurement*, 186, 2021. doi:10.1016/j.measurement.2021.110073.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.