2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

GRANULOMETRY TRANSFORMER: IMAGE-BASED GRANULOMETRY OF CONCRETE AGGREGATE FOR AN AUTOMATED CONCRETE PRODUCTION CONTROL

Max Coenen, Dries Beyer, Michael Haist
Institute of Building Materials Science, Leibniz University Hannover, Germany

Abstract

The size distribution of concrete aggregate can significantly affect the concretes properties (e.g. the consistency and compressive strength). However, the aggregate size distribution can be subject to large fluctuations and is usually unknown during concrete production, leading to a diminished ability of adapting the concrete mix design accordingly. To overcome this limitation, we propose an automatic approach for the estimation of the aggregate size distribution prior to concrete mixing, using image observations of the material on the conveyor belt transporting the aggregate to the concrete mixer. As a result, the derived knowledge about the aggregate size distribution enables a real-time adaptation of the concrete composition for each concrete batch, e.g. by adapting the water demand or the amount of plasticizer accordingly. In particular, we propose *Granulometry Transformer*, a Vision Transformer (ViT) based approach, demonstrating state-of-the-art results on two challenging public benchmark data sets of both, coarse and fine aggregate material.

Introduction

Concrete is one of the most widely used building materials in the world. In total, several billion tons of concrete are produced and used every year. Aggregate, i.e. fine and coarse particles usually of sizes between 0.1 and 32 mm, makes up around 70-80% of the concrete. Due to the large share of aggregate, it significantly influences many important properties - both in the fresh and in the hardened state of the concrete. These include fresh concrete properties such as consistency, workability and segregation tendency as well as hardened concrete properties such as compressive strength, durability, etc. In this context, particularly the size distribution of the aggregates (formally known as grading curve) has a substantial effect on the properties and quality characteristics of the concrete. As a consequence, in order to achieve desired properties (e.g. a target consistency), the aggregate size distribution has to be considered during mix design and concrete production since it significantly affects the water demand or the amount of required superplasticizer. In practice however, the size distribution is usually determined for small samples of the aggregate (a few kilograms) by manual mechanical sieving and is considered representative for a large amount of aggregate (a few tons). Since the aggregate size distribution can exhibit strong variations, especially in cases when e.g. recycled material is used as aggregate, the size distribution of the actual aggregate used for individual production batches of concrete can differ from the predetermined one, and is therefore unknown during production. As a consequence, the mix design is based on imprecise or incorrect assumptions, potentially leading to undesired effects for the final concrete's properties.

In this paper, we propose an approach for the image based prediction of the size distribution of concrete aggregates (cf. Fig. 1), delivering a real-time capable analysis of the size distribution of concrete aggregate. Incorporating such an approach as online measurement process into the production chain of concrete by installing cameras above the aggregate feeding belt allows to derive knowledge about the total amount of aggregates used for the particular concrete batch. As a result, it enables the opportunity to react on detected variations in the size distribution of the aggregate in real-time by adapting the composition, i.e. the mixture design of the concrete accordingly, so that the desired concrete properties are reached. We refer the reader to (Haist et al., 2022), where a concept for the online concrete production control is proposed.

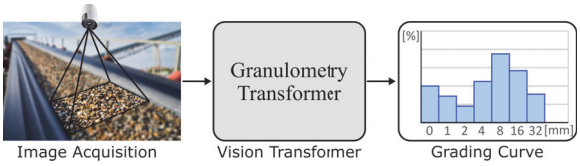


Figure 1: High-level overview on the proposed method.

More specifically, we make the following contributions in this paper.

- We propose a computer vision based strategy for an automated estimation of concrete aggregate size distributions in order to enable a precise concrete production control in real-time.
- We present a deep learning based approach for the determination of aggregate grading curves by building on recent success of transformer networks for vision tasks. In particular, we adapt the architecture of a vision transformer (ViT) network for a patch-based prediction of particle size distributions.
- We perform an extensive evaluation of the proposed framework based on two publicly available and challenging benchmark data sets of concrete aggregate, covering coarse (0-32 mm) and fine (0-2 mm) material. We demonstrate highly promising results obtained by the proposed method, outperforming the quality achieved by a standard ViT architecture.

State of the art

Along with water and cement, aggregate is one of the main basic materials for the production of concrete. As such, the aggregates have a considerable influence on the properties of fresh concrete (e.g. workability and consistency) and hardened concrete (e.g. compressive strength, density, durability). In this context, additionally to technical parameters such as the aggregate density, particle shape, or the chemical composition, it is the particle size distribution which is of particular relevance for concrete production. Knowledge about the aggregate size distribution (also denoted as grading curve) is required for concrete mixture design, since it significantly influences parameters such as water consumption or the demand of plasticizer. In current practice, however, the grading curve of the aggregate is determined for small aggregate samples by mechanical sieving and is considered as representative for many tons of material, neglecting variations and a potentially wide range of material scatter that is typically inherited by aggregates (especially in the case of recycled material). In contrast to current practice based on random sampling, this work proposes an image based strategy for grading curve estimation, allowing to determine the granulometry of the entire material that is actually used for each concrete batch, thus unfolding the potential of a more precise mixture design and enhanced adaptation of the concrete composition (Haist et al., 2022).

The estimation of object size distributions from images has a wide field of interest, ranging from applications in geosciences (Buscombe, 2020), biological and medical applications (Sharma et al., 2020), via hydrological (Chardon et al., 2022) to geographical (Soloy et al., 2020) applications. One line of work towards determining size distributions follows an **object-based procedure**, in which individual objects are segmented first, and their size distribution is computed from the segmentations in a second step. In this context, early approaches for a segmentation based estimation of size distributions were based on grayscale thresholding and morphological operators (Kumara et al., 2012), edge detections (Hamzeloo et al., 2014), or watershed transformations (Lira and Pina, 2006; Terzi, 2017). Modern approaches typically rely on deep learning based methods for the segmentation and granulometry estimation of particles. For example, (Coenen et al., 2021) proposed a convolutional neural network (CNN) based semantic segmentation of concrete aggregate particles and (Soloy et al., 2020) use the Mask RCNN architecture (He et al., 2017) for instance segmenation of pebble grains. A method for the panoptic segmentation of aggregate particles in fresh and hardened concrete was proposed in (Coenen et al., 2022b). However, object based approaches in general suffer from multiple difficulties. On the one hand, they are sensitive to partial occlusions and require an image resolution that is fine enough for recognising the single particles in order to allow the segmentation of individual instances. On the other hand, they demand for an explicit conversion from the two-dimensional segmentations to a volumetric size distribution of the objects (Hamzeloo et al., 2014; Sun et al., 2021), which is only approximate and, therefore, causes inaccuracies for object-centric approaches w.r.t. the task of estimating size distribution.

In contrast to the described object-based procedure, **statistical approaches** avoid the explicit detection and modelling of individual objects by relying on global image statistics in order to predict the size distribution directly from the raw image. In this context, Olivier et al. (2019) and Coenen et al. (2022a) 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 (Olivier et al., 2020; Lang et al., 2021; Sharma et al., 2020).

While the approaches mentioned so far are based on CNN-architectures, recently the application of transformer based models (Vaswani et al., 2017) for vision tasks has shown great potential and encouraging results (Dosovitskiy et al., 2021). In this work, we built upon the recent success of so called Vision Transformers (ViT) and make use of a transformer based architecture for the determination of concrete aggregate size distributions.

Methodology

This paper presents a transformer based approach for the automatic determination of concrete aggregate size distributions. By equipping the conveyor belt in a concrete plant transporting the aggregate material to the concrete mixer with a sensor setup acquiring image sequences of the transported material (cf. Fig. 1), the proposed method can be used to derive detailed knowledge about the actual grading curve of the aggregates used for the production of individual concrete batches in real-time. In this way, an online adjustment of the concrete composition becomes feasible, enabling the real-time control of desired concrete properties as is depicted in Fig. 2 (Haist et al., 2022).

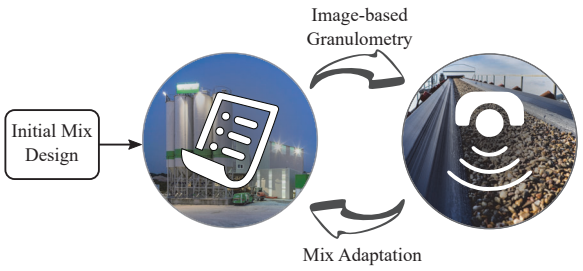


Figure 2: Schematic overview on the concept for a concrete mixture adaptation on the basis of the proposed visual granulometry estimation.

More formally, given an image I depicting concrete aggregates, this paper aims 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 results, 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 op to 1. In order to tackle this problem, we propose the *Granulometry Transformer*, a Vision Transformer based architecture, acting as mapping function $f : I \rightarrow G$.

Overview on Vision Transformers (ViT)

A typical Vision Transformer (ViT) (Dosovitskiy et al., 2021) 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

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

then serves as input to a transformer encoder architecture (Vaswani et al., 2017). As is shown 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. 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) (Vaswani et al., 2017), 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}. As a consequence, the loss for image I can be written as

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

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 such as e.g. the cross-entropy loss for classification problems or the mean squared error for regression problems.

Granulometry Transformer

The typical ViT as just described performs the prediction on image-level based on the embedded cls-token (cf. Eq. 2) and, consequently, neglects the rich information embedded in each of the individual image patches $X_1, ..., X_n$. In this work, we propose to utilise each image patch embedding and to incorporate the patch-level information into the prediction of the particle size distribution. To this end, we decompose the problem of grading curve estimation for the image I into the sub-tasks of estimating the grading curve for small patches extracted from I and, subsequently, aggregate the information to derive the final size distribution for the whole image. More specifically, as is shown in Fig. 3, we make use of the initial ViT-image-decomposition into the sequence of n non-overlapping patches x_i, and predict the grading curve G_i for each individual patch. Note, that in this way, we discard the cls-token z_{cls} and consequently also the cls-embedding X_{cls} from the architectural design of the ViT. Instead, we apply the MLP head H on top of the transformer to each of the final token embeddings X_i (cf. Fig. 3). We make use of the softmax-function as final activation in H, returning $j = 1...N$ output values representing the size distributions G_i which consequently comply with the constraint $\sum p_j = 1$. Under the assumption that each patch shows the same amount (mass) of material, the final size distribution for the whole image is calculated as global average of the patch-wise distributions G_i.

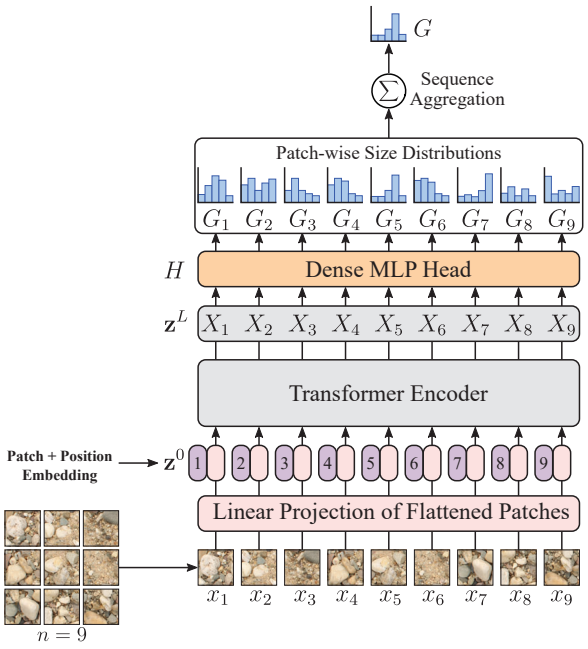


Figure 3: Overview on the proposed Granulometry Transformer architecture. A sequence of linearly projected and flattened image patches are fed to a transformer encoder and a MLP head is used on top of the architecture to predict a particle size distribution for each patch. The final grading curve for the image is calculated by aggregating the patch-wise distributions.

We argue, that by decomposing the task of estimating the

size distribution for an entire image into decoupled tasks of determining the grading curve for smaller individual patches, we are conceptually able to simplify the complexity for the network to learn the mapping between the image and the associated grading curve. In the classical ViT setting using the cls-token for prediction, the transformer encoder has to learn a global latent space embedding as feature representation on image-level on the basis of which the MLP head produces the output. In the case of granulometry estimation, generating a global feature embedding can become challenging since the particle size distribution can locally be highly different, requiring the network to learn a potentially complex aggregation of the local variations into a joint image-level representation. Fig. 4 shows an example of concrete aggregate material which exhibits strong variations in the local size distributions in the image. We believe, by conceptually decomposing the problem into multiple smaller sub-problems, we reduce the complexity of generating the feature embeddings in latent space. By allowing the network to produce patch-wise output distributions that differ from the image-level reference distribution, we enable the network to better account for locally differing appearances, thus relaxing the demand on the network's capability of global reasoning.

Figure 4: Example of concrete aggregate material (left) with locally varying size distributions (right).

Training

According to the proposed architecture shown in Fig. 3, the total image loss for the final grading curve results to

$$L = D\left(\frac{1}{n}\sum_{i=1}^{n} H(X_i),\, Y_I\right). \tag{3}$$

Note that compared to Eq. 2, the proposed loss does not depend on the single class embedding X_{cls}, but instead on all patch embeddings $X_1, \ldots, X_n$, therefore leveraging the additional valuable information encoded in each of the individual tokens. In this paper, we make use of 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

$$D_{\text{KL}} = \sum_{j=1}^{N} p_j \cdot \log\left(\frac{p_j}{\hat{p}_j}\right), \tag{4}$$

where p_j and $\hat{p}_j$ are the reference and the predicted percentiles of the size distribution, respectively. Note that this loss only accounts for errors in the reference bins $p_j > 0$, i.e. in size bins that actually contain material and consequently, an overestimation of empty bins does not contribute to the error metric directly. However, as we treat the grading curve G as a probability distribution with $\sum p_j = 1$, the overestimated values are missing in the bins that are taken into account in the error metric, and therefore, are considered indirectly.

Furthermore, in the setting covered by Eq. 4, the global reference grading curve for the entire image is used for loss computation and thus, training is performed based on a single reference vector as supervision. As a consequence, learning to produce the patch-wise size distributions is only supervised implicitly which, however, allows to apply the proposed method to data for which patch-wise reference information is not available. Nevertheless, we point out that the proposed method allows to perform training in a densely supervised manner in cases where reference distribution for the local patches are available or derivable, e.g. from given instance segmentation masks. As was shown in (Jiang et al., 2021), dense supervision can aid the training process and improve model accuracy.

Experimental Setup

Test data

For the experimental evaluation of the proposed method, we make use of the publicly available *Deep Granulometry Dataset*[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. It is distinguished between the *Coarse Aggregate Data* (D_{coarse}) and the *Fine Aggregate Data* (D_{fine}). Both contain approximately 1700 images associated to one of almost 35 different particle size distributions. While the data in D_{coarse} shows aggregate with particle sizes ranging from 0.1 to 32 mm and provided reference percentiles for $N = 9$ size bins, namely 0.25, 0.5, 1, 2, 4, 8, 16, 32.5, 63 [mm], the D_{fine} data contains 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]. Example images of both data sets showing material with different granularities (from fine to coarse) are shown in Fig. 5.

Test settings

In order to evaluate the grading curve prediction we make use of the described data sets and train networks for each data set 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 an image size of 512x704 [px] for the D_{coarse} data, corresponding to a ground sampling distance (GSD) of 0.5 mm, and to an image size of 480x480 [px] for the D_{fine} data, corresponding to a GSD of 0.1 mm. As network architecture, we adapt the *hybrid ViT-Base*

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

Figure 5: Example images of the two publicly available data sets used for evaluation in this paper.

architecture according to the definition in (Dosovitskiy et al., 2021), i.e. we feed the image to a small CNN backbone and form the input sequence for the transformer based on the feature maps produced by the CNN. More specifically, we apply two convolutional *msEnc-modules* (Coenen et al., 2022a) in case of the D_{coarse} data and one *msEnc-module* in case of the D_{fine} data, which perform residual multi-scale convolutions and downsampling, and which produce feature maps whose size is reduced by factor 0.25 and 0.5 w.r.t. the input image size of the two data sets, respectively. The *ViT-Base* transformer encoder is applied to patches extracted from the feature maps using a patch size of 16x16 [px]. The encoder backbone consists of $L = 12$ layers which are composed of 12 multi-head-self-attention (MSA) modules. Training is done using the Adam optimiser (Kingma and Ba, 2015), a mini-batch size of 24 and an initial learning rate of 10^{-5}. To improve training, the learning rate is decreased by a factor of 10^{-1} after 10 epochs with no improvement in the training loss. To reduce overfitting effects, we apply random radiometric and geometric data augmentations like colour shift, contrast and brightness variations, as well as horizontal and vertical flips.

Evaluation strategy

For the evaluation, we follow a two-fold crossvalidation strategy for both data sets. To this end, we split each data set into three subsets T_1, T_2, and T_3, containing a proportion of 44%, 44%, and 12% of the total amount of images, respectively. To ensure balanced data splits, we divide the data in a way that the proportion of images belonging to the same grading curve is identical across each split. For each data set, we train two networks, alternating between T_1 and T_2 as train and test split, respectively. The T_3 split is used as validation split for both networks. The evaluation in this paper is performed on the joint results obtained by the two networks on both test splits. To this end, we compute the mean absolute errors (MAE) of the predictions for the individual percentiles p_j of the different particle size bins. Furthermore, in order to assess the performance of the grading curve predictions as a total, we make use of the *Hellinger distance* D_{H} (Hellinger, 1909), which is a measure of the similarity between two probability distributions, namely the reference grading curve G and the predicted grading curve $\hat{G}$ in this case. With

$$D_{\text{H}} = \sqrt{1 - BC\left(G, \hat{G}\right)}, \tag{5}$$

where $BC(G, \hat{G})$ is the *Bhattacharyya coefficient* (Bhattacharyya, 1943) defined as

$$BC\left(G, \hat{G}\right) = \sum_{j=1}^{N} \sqrt{p_j \cdot \hat{p}_j}, \tag{6}$$

the distance D_{H} delivers a bounded metric with a maximum distance of 1 and a minimum distance of 0 in case both probability distributions are identical.
In order to compare the results to current state-of-the-art, we report the results on the same data achieved by the *R-S-Net*, a purely CNN-based approach presented by Coenen et al. (2022a), and by a standard Vision Transformer using the ViT-Base architecture proposed in (Dosovitskiy et al., 2021).

Evaluation

In Tab. 1 and 2, the percentile-wise MAE obtained on the D_{coarse} and the D_{fine} data are shown. As is visible from the tables, the results of the *Granulometry Transformer* proposed in this paper for the estimation of particle size distributions are very promising. On the D_{coarse} data, we achieve percentile-wise MAEs between 0.19 % and 1.81 % resulting in an average MAE of 1.08 %. On the D_{fine} data set, the values for the MAEs are comparably larger, ranging between 3.19 % and 4.71 %, with an average MAE of 3.79 %. The same behaviour is observable also for the results obtained by the R-S-net and the standard ViT. We identify two potential reasons for the differences in the performance on the two data sets. One reason might be the different numbers of percentiles N that are differentiated by the two data sets. While the D_{coarse} data considers nine grain size bins, the number of distinguished percentiles of the D_{fine} data is only six. Because the softmax-activation, used as normalisation by the final layer of the networks, constraints the output to sum up to 1, the total magnitude of the prediction errors is distributed over the amount of percentiles, leading to decreased percentile-wise average errors the more bins are considered for estimation. The second reason might be related to the GSD of the data in combination with the particle sizes covered by the data sets. While the coarse aggregate data contains particles with a maximum size of 32 mm and is used with a GSD of 0.5 mm, leading to an image representation of the largest grain by 64 px, the fine aggregate data containing a maximum grain size of 2 mm is used with a GSD of 0.1 mm, leading to an representation of the maximum particle in the image with a size of only 20 px. As a consequence, it is possible, that the fewer amount of image

Table 1: Quantitative results obtained on the $\mathbf{D}_{coarse}$ test data. The table shows the percentile-wise MAE and their average (Ø) in [%].

Grain size bins [mm]	**0.25**	**0.5**	**1**	**2**	**4**	**8**	**16**	**31.5**	**63**	**Ø**
R-S-Net (Coenen et al., 2022a)	0.19	1.15	1.27	0.62	1.40	1.51	1.62	1.72	0.20	1.08
ViT (Dosovitskiy et al., 2021)	0.21	1.63	1.77	0.93	1.24	2.80	2.42	2.94	0.21	1.57
Ours	0.19	1.37	1.42	0.69	1.12	1.81	1.46	1.43	0.21	1.08

Table 2: Quantitative results obtained on the $\mathbf{D}_{fine}$ test data. The table shows the percentile-wise MAE and their average (Ø) in [%].

Grain size bins [mm]	**0.063**	**0.125**	**0.25**	**0.5**	**1**	**2**	**Ø**
R-S-Net (Coenen et al., 2022a)	4.25	3.74	4.03	2.81	2.42	3.77	3.50
ViT (Dosovitskiy et al., 2021)	4.26	4.11	4.47	3.89	4.02	4.51	4.21
Ours	3.83	3.59	4.10	3.29	3.19	4.71	3.79

information per particle of the D_{fine} data in comparison to the D_{coarse} data can cause a decreased performance in estimating the particle size distribution for the fine data set.

In addition to the MAEs, Tab. 3 contains the average *Hellinger distances* obtained by the different methods on the two aggregate data sets. As can be seen from the average MAE (cf. Tab. 1 and Tab. 2) as well as from the mean Hellinger distances (Tab. 3), the proposed *Granulometry Transformer* performs significantly better on both data sets compared to the standard ViT used as baseline. Regarding the average MAE, our method yields errors which are smaller by a difference of 0.49 % and 0.42 % on the D_{coarse} and D_{fine} data set, respectively, which correspond to an relative improvement of approximately 30 %. Similar relative improvements are obtained for the mean Hellinger distance on the D_{coarse} data, while the relative improvements on the D_{fine} data are with approximately 12 % comparably smaller. Compared to the R-S-Net used as purely convolutional baseline, our approach performs on-par w.r.t. both metrics, the MAEs and the Hellinger distances.

Table 3: Average Hellinger distances obtained on the two data sets D_{coarse} and D_{fine}.

Mean Hellinger distances	D_{coarse}	D_{fine}
R-S-Net (Coenen et al., 2022a)	0.500	0.101
ViT (Dosovitskiy et al., 2021)	0.071	0.127
Ours	0.048	0.112

To obtain more detailed insights into the distribution of the percentile-wise absolute errors and the Hellinger distances, Fig. 6 and Fig. 7 show the cumulative histogram of both metrics for the two test data sets, respectively. Again, the depicted graphs highlight the previous observation of our approach performing on-par with the CNN-based approach and outperforming the ViT baseline. Also, the performance differences between both data sets become visible by the cumulative histograms exhibiting a steep incline on the D_{coarse} data and a comparably flatter ascent of the curve on the D_{fine} data set.
Since the data sets consists of multiple images associated to identical particle size distributions (approx. 50 images per grading curve), it is possible to agglomerate all predictions of images having the same reference grading curve by computing the average predicted value and standard

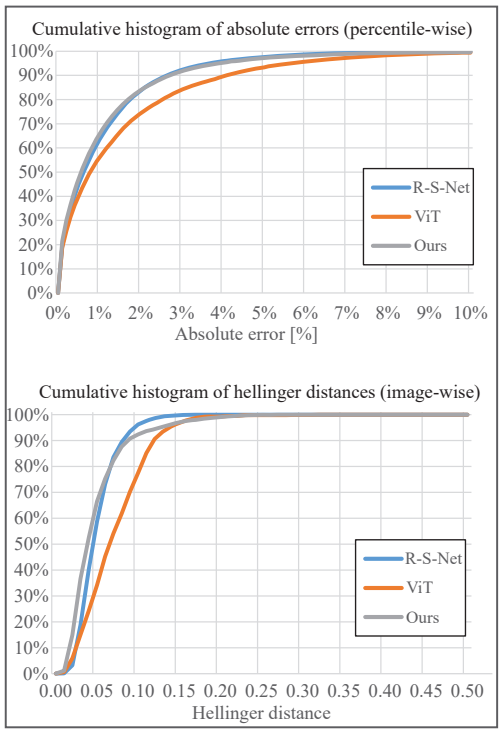


Figure 6: Cumulative histogram of absolute errors (top) and of the Hellinger distance (bottom) obtained on the D_{coarse} data set.

deviation for each percentile for a visual demonstration. Fig. 8 shows representative examples of aggregate images and their associated grading curves of both, the reference (green) as well as the average prediction (red). Furthermore, the bands depicted in the diagrams represent the standard deviation area of predictions of the test images of the respective aggregate sample.

Conclusion

The current practice of concrete mixture design and production often relies on strong assumptions on properties of the raw materials used for production, such as e.g. the particle size distribution of the aggregate. As a consequence, unknown variations of the aggregate's grading curve and deviations from the assumed size distribution (especially pronounced in the case of recycled materials), are not properly taken into account in the mix

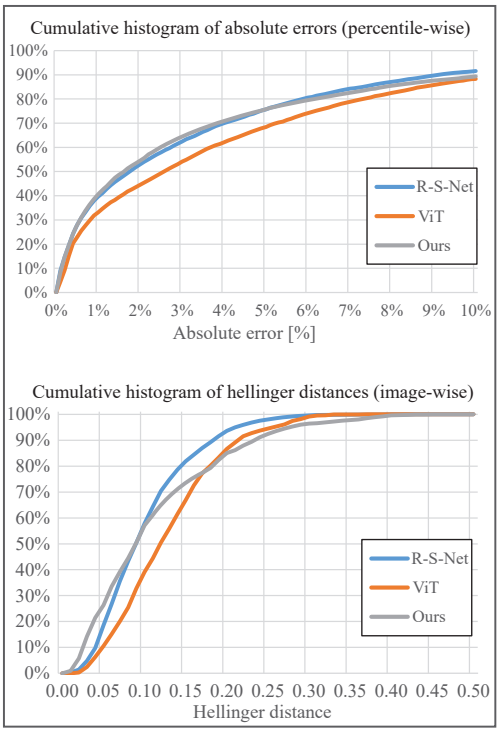
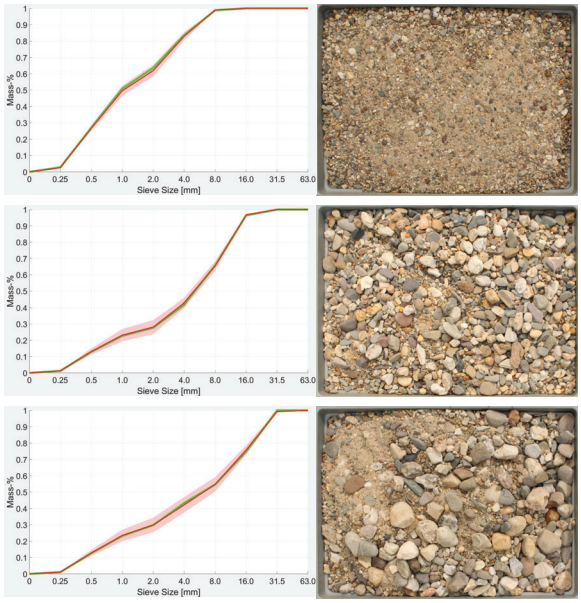


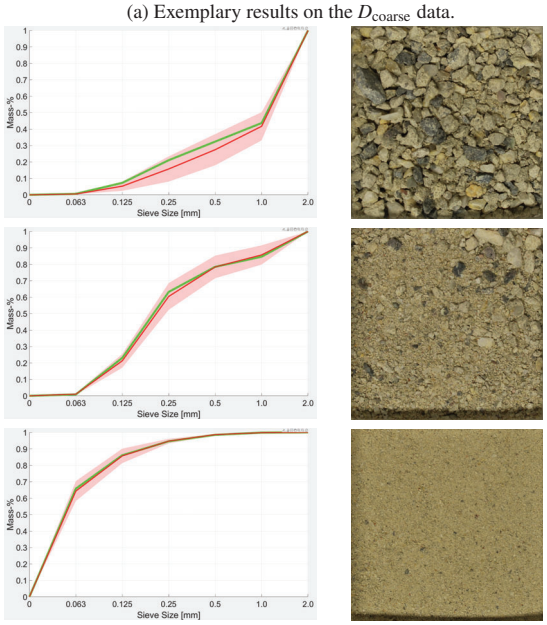
Figure 7: Cumulative histogram of absolute errors (top) and of the Hellinger distance (bottom) obtained on the D_{fine} data set.

design, potentially leading to undesired effects w.r.t. the concrete properties. In order to enable the control of the variations, we present a deep learning based method for the determination of the grading curve of concrete aggregate from images. More specifically, we propose an adaptation of the Vision Transformer architecture for the local patch-aware determination of the particle size distribution, leading to an improved performance compared to the standard ViT network. Experiments on two challenging data sets demonstrated highly promising results with percentile-wise mean average errors of less than 2 % and 5 % obtained on both data sets, respectively.

In the future, we aim at extending the approach by not only determining the particle size distribution of concrete aggregate, but also at identifying shape-related properties of the aggregate and at predicting its material composition (e.g. the material composition of recycled aggregate), two properties which carry highly relevant information regarding the concrete mix design. Furthermore, we aim at applying the proposed approach as a basis for the development of an online concrete control scheme, with the goal to adapt the mix composition in real time to react to the detected fluctuations and properties of the raw materials. This implies the extension of the described approach from working in a static scenario (single images are processed) to an application in a dynamic scenario (image sequences obtained from the material moving on the conveyor belt).



(a) Exemplary results on the D_{coarse} data.



(b) Exemplary results on the D_{fine} data..

Figure 8: Examples of aggregate images (right) belonging to different reference grading curves (left, green) and the average predicted size distribution (red curve) as well as its standard deviation (light red area).

Acknowledgements

The work is part of the project *ReCyCONtrol* funded by the German Federal Ministry of Education and Research (BMBF) under the grant No. 033R260A.

References

Bhattacharyya, A. (1943). On a Measure of Divergence between two Statistical Populations defined by their Probability Distributions. Bulletin of the Calcutta Mathemat-

ical Society, 35:99–109.

Buscombe, D. (2020). SediNet: A configurable Deep Learning Model for mixed qualitative and quantitative optical Granulometry. Earth Surface Processes and Landforms, 45(3):638–651.

Chardon, V., Piasny, G., and Schmitt, L. (2022). Comparison of Software Accuracy to estimate the Bed Grain Size Distribution from digital Images: A Test performed along the Rhine River. River Research and Applications, 38(2):358–367.

Coenen, M., Beyer, D., Heipke, C., and Haist, M. (2022a). Learning to Sieve: Prediction of Grading Curves from Images of Concrete Aggregate. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, volume V-2-2022, pages 227–235.

Coenen, M., Schack, T., Beyer, D., Heipke, C., and Haist, M. (2021). Semi-Supervised Segmentation of Concrete Aggregate Using Consensus Regularisation and Prior Guidance. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, volume V-2-2021, pages 83–91.

Coenen, M., Schack, T., Beyer, D., Heipke, C., and Haist, M. (2022b). ConsInstancy: Learning Instance Representations for Semi-Supervised Panoptic Segmentation of Concrete Aggregate Particles. Machine Vision and Applications, 33(57).

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In International Conference on Learning Representations (ICLR).

Haist, M., Heipke, C., Beyer, D., Coenen, M., Vogel, C., Schack, T., Ponick, A., and Langer, A. (2022). Digitization of the Concrete Production Chain using Computer Vision and Artificial Intelligence. In Proceedings of the 6th fib Congress, pages 434–443.

Hamzeloo, E., Massinaei, M., and Mehrshad, N. (2014). Estimation of Particle Size Distribution on an Industrial Conveyor Belt using Image Analysis and Neural Networks. Powder Technology, 261:185–190.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In IEEE International Conference on Computer Vision (ICCV), pages 2980–2988.

Hellinger, E. (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. Journal für die reine und angewandte Mathematik, 136:210–271.

Jiang, Z.-H., Hou, Q., Yuan, L., Zhou, D., Shi, Y., Jin, X., Wang, A., and Feng, J. (2021). All Tokens Matter: Token Labeling for Training better Vision Transformers. In Advances in Neural Information Processing Systems (NeurIPS), volume 34, pages 18590–18602.

Kingma, D. and Ba, L. (2015). Adam: A Method for Stochastic Optimization. In International Conference on Learning Representations (ICLR).

Kumara, G. H. A. J. J., Hayano, K., and Ogiwara, K. (2012). Image Analysis Techniques on Evaluation of Particle Size Distribution of Gravel. International Journal of Geomate, 3:290–297.

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

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

Olivier, L. E., Maritz, M. G., and Craig, I. K. (2019). Deep Convolutional Neural Network for Mill Feed Size Characterization. IFAC-PapersOnLine, 52(14):105–110.

Olivier, L. E., Maritz, M. G., and Craig, I. K. (2020). Estimating Ore Particle Size Distribution using a Deep Convolutional Neural Network. IFAC-PapersOnLine, 53(2):12038–12043.

Sharma, K., Gold, M., Zurbruegg, C., Leal-Taixe, L., and Wegner, J. D. (2020). HistoNet: Predicting Size Histograms of Object Instances. In IEEE Winter Conference on Applications of Computer Vision (WACV), pages 3637–3645.

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

Sun, Z., Liu, H., Huyan, J., Li, W., Guo, M., Hao, X., and Pei, L. (2021). Assessment of Importance-based Machine Learning Feature Selection Methods for Aggregate Size Distribution Measurement in a 3D binocular Vision System. Construction and Building Materials, 306.

Terzi, M. (2017). Particle Size Distribution Analysis in Aggregate Processing Plants using digital Image Processing Methods. Romanian Journal of Materials, 47(4):514–521.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is All you Need. In Advances in Neural Information Processing Systems (NIPS), volume 30.