Computers and Concrete, *Vol. 25, No. 5* (*2020*) *433-445*
DOI: https://doi.org/10.12989/cac.2020.25.5.433

Estimating the workability of self-compacting concrete in different mixing conditions based on deep learning

Liu Yang and Xuehui An*

State Key Laboratory of Hydro Science and Engineering, Tsinghua University, 100084-Beijing, China

(*Received October 21, 2019, Revised March 31, 2020, Accepted April 10, 2020*)

Abstract. A method is proposed in this paper to estimate the workability of self-compacting concrete (SCC) in different mixing conditions with different mixers and mixing volumes by recording the mixing process based on deep learning (DL). The SCC mixing videos were transformed into a series of image sequences to fit the DL model to predict the SF and VF values of SCC, with four groups in total and approximately thirty thousand image sequence samples. The workability of three groups SCC whose mixing conditions were learned by the DL model, was estimated. One additionally collected group of the SCC whose mixing condition was not learned, was also predicted. The results indicate that whether the SCC mixing condition is included in the training set and learned by the model, the trained model can estimate SCC with different workability effectively at the same time. Our goal to estimate SCC workability in different mixing conditions is achieved.

Keywords: deep learning; self-compacting concrete; workability; mixing condition; mixer; mixing volume; slump flow and V-funnel test; convolutional neural network; recurrent neural network

1. Introduction

Self-compacting concrete (SCC) can flow to fill gaps in reinforcements, corners of moulds, and voids in rock blocks without vibration during the placement process (Okamura and Ouchi 2003, An *et al*. 2014). SCC of high quality has superior workability, including high deformability and segregation resistance. The workability determines whether a concrete is qualified for use. Therefore, evaluating the workability of SCC before placement is of incredible significance. Slump flow and V-funnel tests (Okamura and Ouchi 2003, An *et al*. 2014, Wu and An 2014) are the most commonly used tests in estimating SCC workability. The slump flow value (SF) indicates the SCC deformability, and the V-funnel flow time (VF) indicates the segregation resistance ability and viscosity. Therefore, the two tests are always performed for mix design in laboratories and for quality control on construction sites. However, these tests have a number of problems in practice. One is that these tests are always conducted before placement to ensure that the SCC has good workability. In addition, the workability of SCC can only be determined immediately before placement. If the concrete is unqualified, the mixture is wasted, and the mix design needs to be changed (Gidaris *et al*. 2015).

The issues mentioned above can be addressed if the SCC workability can be estimated during the mixing process. Chopin *et al*. (2017) proposed a method that a concrete mixture can be regarded as a large rheometer. The rheological parameters of SCC are determined by considering fresh concrete as a Bingham material. Beaupré (1994) indicated that the rheological parameters can be used to estimate whether a mixture is a qualified SCC. However, it remains an open question to use a mixer as a reliable rheometer. Other methods have been proposed to evaluate SCC workability. Some experienced engineers can evaluate the concrete by watching the concrete mixing process. Thus, it may be an effective solution to use visual information, such as in image processing methods (Marinoni *et al*. 2005, Hutchinson and Chen 2006, Cabaret *et al*. 2007, Lee *et al*. 2013), during the mixing process to predict SCC workability. Li and An (2014) adopted this method and found a relationship between the visual characteristics and the workability of SCC. Daumann and Nirschl (2008) used ultramarine blue as a tracer component to obtain the mixture homogeneity during the concrete mixing process. However, these image processing methods depend mainly on human experience and insight. For example, in Li's work, the boundaries were manually extracted. In Daumann's work, features such as the tracer component were also manually selected. In addition, these features were often connected to a specific experimental scenario. Consequently, the application of these methods is limited.

The aforementioned methods contain several drawbacks. To solve these problems, it is necessary to reduce the dependency on human experience and investigate the information hidden behind the original data. Deep learning (DL) could be an effective alternative method to automatically estimate SCC workability. DL solves the data representation problem by introducing relatively simple intermediate representations that can be combined to establish complex concepts. Thus, there is no need for specific techniques to extract features that can represent the image data (Deng *et al*. 2009, Bengio *et al*.

*Corresponding author, Professor
E-mail: anxue@tsinghua.edu.cn

ISSN: 1598-8198 (Print), 1598-818X (Online)

2013, He *et al*. 2014). Due to its complex structure, DL requires a large amount of data to generate models with good performance in prediction and therefore has a large computational cost. Deep convolutional neural networks (CNNs) are an example of a successful DL model that provides a powerful tool for extracting visual representations and is widely used in image classification (Affonso *et al*. 2017), object detection (Girshick *et al*. 2014), and semantic segmentation (Gidaris *et al*. 2015). Compared with CNNs, recurrent neural network (RNN) models are "deep in time" and can form implicit compositional representations in the time domain (Donahue *et al*. 2017). The obvious drawback of RNNs is the vanishing gradient effect, which refers to the tendency to back propagate an error signal through a long-range temporal interval. This problem becomes increasingly difficult to solve. Long short-term memory (LSTM) units are recurrent modules that enable long-range learning. LSTM units have a hidden state augmented with nonlinear mechanisms that allow state update, reset and propagation without modification using simple learned gating functions. LSTMs have been used for sequential labelling and provide much improvement when enough data are available. It has been demonstrated that a CNN with RNN units can be used for visual time-series modelling (Donahue *et al*. 2017).

Machine learning techniques have been successfully employed in construction engineering in recent years. This approach can automatically establish a relationship between the original information and the parameter of interest based on a well-trained model. Machine learning has been used to detect cracks (Yokoyama and Matsumoto 2017, Zhang *et al*. 2017), predict compressive strength (Chou *et al*. 2014, Abd and Abd 2017, Yaseen *et al*. 2018), analyze concrete dam reliability (Hariri-Ardebili and Pourkamali-Anaraki 2018), assess durability (Taffese and Sistonen 2017), and predict carbonation (Taffese *et al*. 2015). Ding and An (2018) proposed an approach that uses a DL model to automatically extract features from mixing images and predict the SF and VF of SCC. However, this approach proves only that the DL model can be used for a specific mixer and mixing volume, constituting merely one mixing condition. There are various kinds of SCC with different mixers and mixing volumes whose workability needs to be predicted in reality. The mixing conditions are diverse and complicated. Whether DL method can be generalized over a larger mixing scope, namely, in different kinds of mixers mixing different SCC volumes, is important and needs to be demonstrated. Additionally, if this could be realized, the mixing conditions including mixers and mixing volumes were countless in the world though. It is logical to collect all the mixing conditions to train and predict. However, this work would cost a lot and hard to realize. Is it possible to choose some typical mixing conditions to train and predict other new mixing conditions? In this way, some mixing conditions do not have to be collected because they can be predicted by others. Therefore, the workload can be reduced and made feasible. This idea needs to be proven too.

In all, there were two main objectives in this study. One was to extend the use of the DL model to estimate the workability of SCC in different mixing conditions, including different types of mixers and mixture volumes. The other was to demonstrate that a number of typical mixing conditions combined together can be used to predict a new mixing condition.

Four groups of SCC mixing videos with different mixing conditions were collected. Each group contained many SCC mixing videos with the same mixing conditions but disperse workability. All videos were processed into image sequences in a suitable format for DL. The data were then divided into the training and validation set and testing sets, with about 30 thousand samples with SF and VF as its label. The testing sets were divided into two kinds according to whether their mixing conditions (including the mixer and mixing volume) were contained in the training sets. Then, the model was trained to find relationships between the image sequences and SCC workability. After that, two kinds of testing data distinguished by whether their mixing conditions were learned by the training sets were predicted to obtain the SF and VF. Based on this approach, a method to evaluate SCC workability in different mixing conditions altogether at the same time was proposed, and the goal to perform effective prediction on a new mixing condition was achieved.

2. Data processing

DL requires a large amount of data to generate models with good performance in prediction. Therefore, the main purpose of this part is to collect and prepare enough data suitable for the DL model. Among these data, preprocessing method was applied to overcome the overfitting problem, which is typical in DL field and effects the prediction accuracy a lot. Data augmentation was further used to reduce the problem and provide enough suitable data.

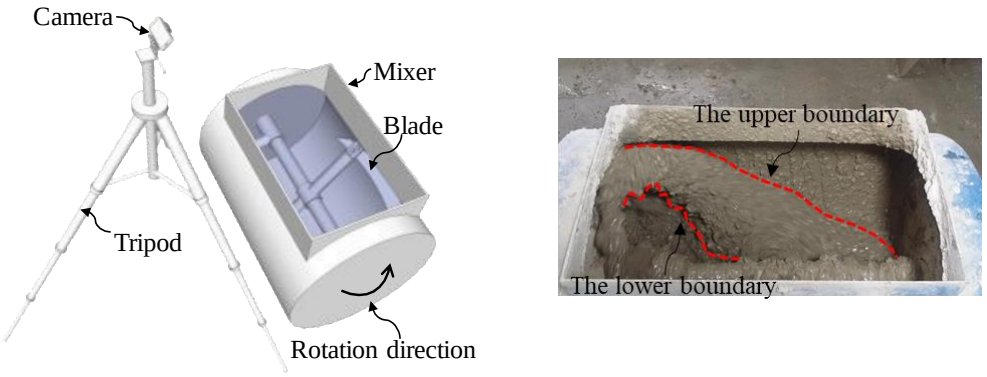
2.1 Data collection

The DL model takes image sequences of the SCC mixing process as input. The image sequences were collected by recording videos of the SCC mixing process in four mixing conditions, namely, groups A, B, C and D. The SCC of group A was mixed in a 30 L single-shaft mixer with a 20 L mixing volume. The mixer of groups B, C and D is a 60 L single-shaft mixer. The mixing volumes of group B, group C and group D were 15 L, 40 L, and 20 L, respectively shown in Table 1.

The experimental setup to record videos is shown in Fig. 1(a). A camera on a tripod was placed beside the mixer. The shooting angle and tripod position in different experiment batches were adjusted to record the two important boundaries of the mixing SCC as shown in Fig. 1 because

Table 1 The mixing parameters of the four groups

Group name	Mixer volume (L)	Mixing volume (L)	The Number of mixing videos
Group A	30	20	31
Group B	60	15	21
Group C	60	40	36
Group D	60	20	6



(a) The relative placement of the camera and mixer (b) Two important boundaries of SCC

Fig. 1 Schematic diagram of the test setup

Table 2 Mix composition and workability characteristics of the mixing videos in group A

Video number	*Vw/Vc*	SP%	Cement (kg/m^3)	Water (kg/m^3)	Plasticizer (kg/m^3)	Hold time (min)	SF (mm)	VF (s)	Volume (L)
1	1.15	0.85	515	188	4.38	0	600	13.0	20
2	1.20	1.00	504	191	5.04	0	610	9.9	20
3	1.30	0.90	482	199	4.34	0	610	18.1	20
4	1.15	0.90	515	188	4.64	0	648	13.0	20
5	1.13	0.85	520	186	4.42	0	598	16.4	20
6	1.11	0.85	525	185	4.46	0	625	26.0	20
7	1.17	0.90	511	189	4.60	0	668	9.8	20
8	1.11	0.90	525	184	4.73	0	675	13.0	20
9	1.13	0.90	520	186	4.68	0	675	13.0	20
10	1.17	0.85	511	189	4.34	0	680	11.6	20
11	1.30	1.00	482	198	4.82	0	690	9.9	20
12	1.30	1.00	482	198	4.82	0	700	10.2	20
13	1.20	0.80	504	192	4.03	30	350	24.1	20
14	1.20	0.80	504	192	4.03	0	480	14.5	20
15	1.20	0.90	504	191	4.53	60	510	14.7	20
16	1.20	1.00	504	191	5.04	30	585	9.2	20
17	1.10	1.00	528	183	5.28	0	460	45.5	20
18	1.10	0.90	528	183	4.75	30	355	*Blocked*	20
19	1.20	1.00	504	191	5.04	90	360	*Blocked*	20
20	1.30	0.90	482	199	4.34	30	485	69.0	20
21	1.30	0.90	482	199	4.34	60	340	*Blocked*	20
22	1.30	1.00	482	198	4.82	90	440	*Blocked*	20
23	1.20	1.00	504	191	5.04	60	495	56.8	20
24	1.20	0.90	504	191	4.53	30	565	83.0	20
25	1.10	0.90	528	183	4.75	0	575	35.8	20
26	1.20	0.90	504	191	4.53	0	705	43.9	20
27	1.30	1.00	482	198	4.82	60	655	30.8	20
28	1.18	0.85	508	190	4.32	0	668	64.0	20
29	1.15	0.88	515	188	4.54	0	679	31.0	20
30	1.30	1.00	482	198	4.82	30	690	52.9	20
31	1.20	0.90	504	191	4.53	0	710	52.0	20

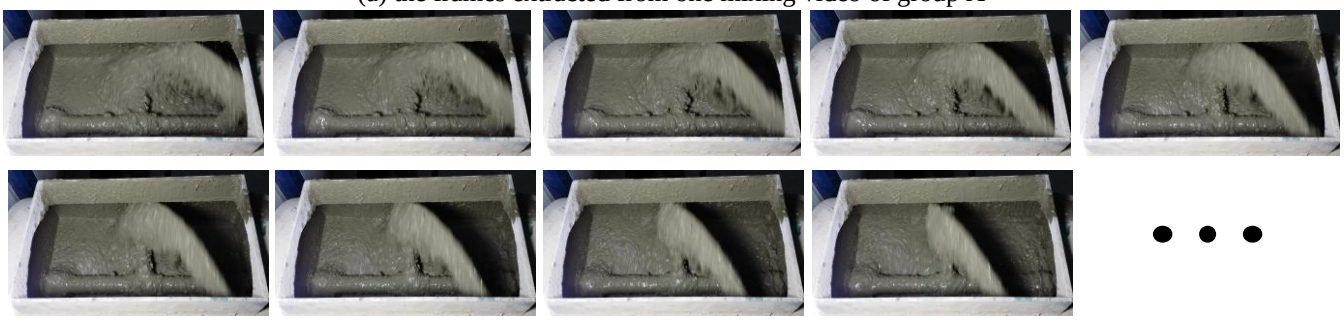
the upper boundary and lower boundary have proven to be the most prominent feature to distinguish SCC with different workability (Li and An 2014). Different tripod location and shooting angles introduced different perspective distortions into the videos. The camera was a smartphone with a frame rate of 30 fps. The resolution of each picture was fixed at 1920 pixels×1080 pixels. After mixing and recording, slump flow and V-funnel tests were performed to obtain the SF and VF values as a label and name.

The cement used in all experiments was 42.5 Portland cement. The coarse aggregates contained crushed stones with particle sizes from 5 mm to 20 mm. The fine aggregates were quartz sands with a maximum particle size of 5 mm. Polycarboxylate super plasticizer (SP) was used as a water reducing agent with a 20% solid content. Fly ash

Table 3 Mix composition and workability characteristics of the mixing videos in group B

Video number	*Vw/Vc*	SP%	Cement (kg/m^3)	Fly ash (kg/m^3)	Water (kg/m^3)	Plasticizer (kg/m^3)	SF (mm)	VF (s)	Volume (L)
1	1.40	1.20	474	0	211	5.69	600	7.8	15
2	1.30	0.80	517	69	201	7.41	638	14.8	15
3	1.07	1.43	455	0	217	5.46	675	28.3	15
4	1.50	1.20	412	176	201	5.88	700	9.8	15
5	1.04	1.43	494	0	205	4.94	605	36.1	15
6	1.30	1.00	455	0	218	5.00	625	32.7	15
7	1.50	1.10	474	0	212	4.25	645	34.8	15
8	1.40	0.90	474	0	212	4.74	650	34.7	15
9	1.40	1.00	474	0	211	6.16	660	44.5	15
10	1.40	1.30	517	0	198	4.14	430	160.9	15
11	1.20	0.80	517	0	198	4.65	535	100.8	15
12	1.20	0.90	494	0	205	4.45	540	54.5	15
13	1.30	0.90	494	0	206	3.96	555	65.7	15
14	1.30	0.80	517	0	197	5.17	565	56.0	15
15	1.20	1.00	474	0	212	3.79	570	42.8	15
16	1.40	0.80	494	0	204	5.44	433	15.6	15
17	1.30	1.10	362	151	198	3.51	495	16.0	15
18	1.16	0.97	494	0	205	4.94	485	15.3	15
19	1.30	1.00	494	0	204	5.93	523	13.3	15
20	1.30	1.20	474	0	212	4.74	535	10.8	15
21	1.40	1.00	474	0	211	5.44	568	26.9	15

(a) the frames extracted from one mixing video of group A



(b) the frames extracted from one mixing video of group C

Fig. 2 The frames extracted from mixing videos

was used as an addictive in group B.

During the mixing and recording process, the camera was used to record videos through the opening hatch of the mixer. First, all the dry materials, including cement, sand and gravel, were mixed for 30 s, and then water and SP were added. The materials were continuously wet-mixed for 240 s. When the mixing ended, the mixture was poured out, followed by slump flow and V-funnel tests to obtain the SF and VF values of SCC. Sometimes, the produced SCC was put in a container for a certain time (such as 30 min and 60 min) as a hold time and then poured back into the mixer for remixing. The slump flow and V-funnel tests were then performed again. Videos were also collected during these remixing tests.

For group A, the SCC was mixed in a 30 L single-shaft mixer with a 20 L mixing volume. Thirty-one videos with different workability characteristics are listed in Table 2. They are distinguished from each other due to different

Table 4 Mix design and workability properties of the SCC in the group C videos

Video number	*Vw/Vc*	SP%	Cement (kg/m^3)	Water (kg/m^3)	Plasticizer (kg/m^3)	Hold time (min)	SF (mm)	VF (s)	Volume (L)
1	1.05	0.95	566	187	5.37	0	660	19.3	40
2	1.14	1.51	511	182	7.71	0	685	14.4	40
3	1.09	1.00	511	175	5.11	0	695	14.2	40
4	0.98	1.30	616	187	7.97	0	705	17.9	40
5	1.13	0.95	610	218	5.80	0	605	5.2	40
6	1.07	1.29	511	172	6.61	0	665	21.4	40
7	1.07	1.18	511	172	6.04	0	685	9.0	40
8	1.09	1.48	511	174	7.55	0	690	17.3	40
9	1.06	1.27	511	169	6.48	0	703	18.2	40
10	1.05	1.06	566	187	5.97	0	720	18.0	40
11	1.11	1.49	511	177	7.61	20	720	17.8	40
12	1.00	1.32	603	187	7.97	25	740	18.1	40
13	1.32	0.89	592	248	5.27	0	365	7.0	40
14	1.32	0.89	592	248	5.27	0	425	4.7	40
15	1.19	0.89	592	223	5.27	0	350	12.2	40
16	1.03	0.95	610	198	5.80	0	548	9.3	40
17	1.02	0.89	627	202	5.58	0	350	16.7	40
18	1.08	0.89	627	215	5.58	0	385	7.0	40
19	1.58	0.89	592	298	5.27	0	438	2.5	40
20	1.06	1.22	511	169	6.26	0	495	35.0	40
21	1.06	1.27	511	169	6.51	0	500	44.0	40
22	1.05	1.10	511	169	5.61	30	420	56.4	40
23	1.05	0.95	511	169	4.86	0	405	*Blocked*	40
24	0.99	0.99	603	187	5.97	0	395	33.6	40
25	1.01	1.54	561	175	8.66	0	425	72.8	40
26	0.90	0.95	610	173	5.80	0	433	*Blocked*	40
27	1.05	1.13	511	169	5.76	0	440	40.5	40
28	1.09	1.38	511	174	7.03	0	550	30.8	40
29	0.97	0.89	627	192	5.58	0	300	*Blocked*	40
30	1.11	0.89	592	208	5.27	0	300	*Blocked*	40
31	1.11	1.69	511	175	8.66	0	708	66.9	40
32	1.09	1.58	511	174	8.05	0	650	63.2	40
33	1.08	1.40	511	172	7.18	0	665	104.2	40
34	1.11	1.49	511	177	7.61	0	683	100.5	40
35	1.09	1.13	511	175	5.79	0	675	68.2	40
36	1.00	1.32	603	187	7.97	0	735	60.1	40

water to cement ratios by volume (Vw/Vc), SP contents of the cement by weight (SP%), and hold times. The SCC formulations in video 18, 19, 21, 22 became blocked in the V-funnel test due to their large viscosities.

For group B, the SCC was mixed in a 60 L single-shaft mixer with a 15 L mixing volume. A total of twenty-one mixing videos are listed in Table 3.

For group C, the mixing volume of the SCC was the largest, at 40 L. The SCC was mixed in a 60 L single-shaft mixer. Thirty-six videos were collected, as shown in Table 4. As noted, the mixtures in videos 23, 26, 29, and 30 were blocked. For group C, the mixing volume was relatively large, at 40 L. To save time as well as materials to collect more videos, after conducting the slump flow and V-funnel tests, the SCC was put back to the mixer, and then water or SP were added and the SCC was remixed to repeat the workability tests to obtain new SF and VF values.

For group D, the SCC was mixed in a 60 L single-shaft mixer with a 20 L volume whose volume was between that of group B (15 L) and group C (40 L). Six videos were collected, as shown in Table 5.

Four groups of videos with different mixers and mixing volumes were gathered. Their parameters are presented in Table 1.

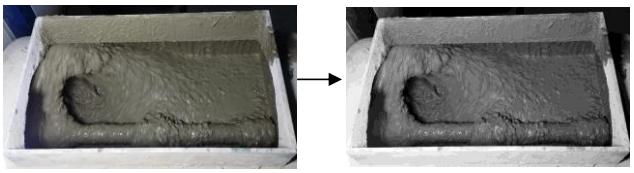
According to the process of data collection, the necessary spatial and temporal information was included in the video data for the DL model to learn. The SF and VF values were used as two labels of the corresponding SCC mixing video, composing a vector with two elements. The VF value of the blocked V-funnel test was set to a default value of 200 to satisfy the need for numerical labels, which was large enough compared with other cases. After recording, the videos were converted into numerous images with SF and VF as its label, shown in Fig. 2, and prepared for the input of the DL model.

2.2 Data preprocessing

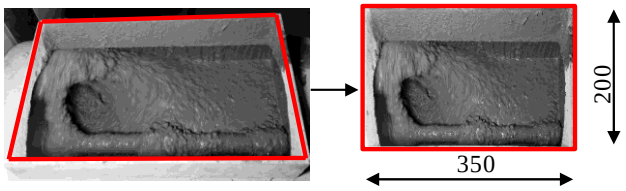
After the data was collected and transformed into images, they cannot be directly input into the DL model for

Table 5 Mix composition and workability characteristics of the mixing videos in group D

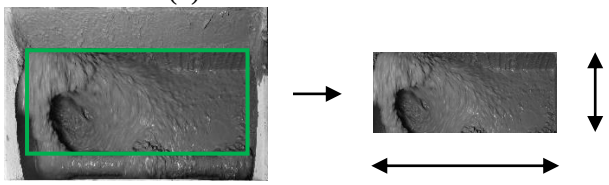
Video number	*Vw/Vc*	SP%	Cement (kg/m^3)	Fly ash (kg/m^3)	Water (kg/m^3)	Plasticizer (kg/m^3)	SF (mm)	VF (s)	Volume (L)
1	1.16	0.99	362	151	198	3.60	655	68.4	20
2	1.15	0.93	310	201	198	2.88	660	22.5	20
3	1.15	1.07	310	201	198	3.32	550	30.7	20
4	1.15	1.07	310	201	198	3.32	660	55.2	20
5	1.15	0.88	310	201	198	2.73	455	21.2	20
6	1.15	1.07	310	201	198	3.32	465	17.6	20



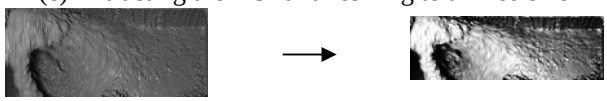
(a) Converting the RGB image into grayscale



(b) Affine transformation

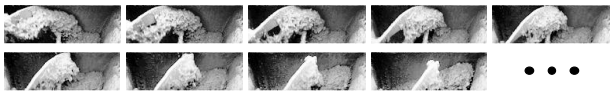


(c) Extracting the ROI and resizing to a fixed size

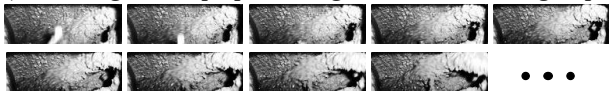


(d) Histogram equalization

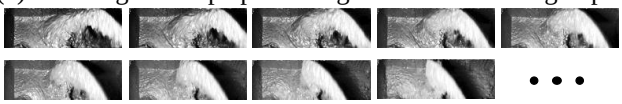
Fig. 3 Date preprocessing



(a) The images after preprocessing from one video of group A



(b) The images after preprocessing from one video of group B



(c) The images after preprocessing from one video of group C

Fig. 4 The images after preprocessing in several groups

training and validation. These images require preprocessing due to two reasons. First, the size of the original image is quite large, 1920 pixels×1080 pixels, placing a significant demand on the computational requirements of the neural networks. In addition, not all the information is necessary, and some information may be noise and may severely impact the performance of the DL model in the form of overfitting. The overfitting problem occurs when a model fits the training data too well. The details and noise in the training data are learned by the model to such an extent that it negatively impacts the performance of the model on new data. Noise or random fluctuations in the training data are picked up and mistakenly learnt as features by the model. However, as these concepts do not apply among new data, it negatively affects the model's ability to provide accurate predictions. Therefore, the preprocessing is necessary for every image before training.

The preprocessing procedure comprises four main steps as shown in Fig. 3. First, the RGB images were converted into grayscale. Second, the affine transformation was carried out. Third, the region of interest (ROI) was extracted. Finally, histogram equalization was performed as presented and resized the image to a fixed size.

The RGB image has three channels, including R (red), G (green) and B (blue), while the grayscale image has only one channel, so the computational cost can be reduced by two times, increasing the computing speed. In addition, the colour of the SCC is different because of the shooting environment and camera parameters, so the colour provides little information to differentiate the SCC workability. Thus, the RGB image was converted into grayscale. The effect of this operation is shown in Fig. 3(a).

Different shooting angles and camera placements result in differences in the perspective distortions among videos. However, the differences in SCC workability is not essentially due to these distortions. Taking distortions as a feature would confuse the computer and affect its performance on new data. The affine transformation was carried out to correct for nonideal camera angles and eliminate distortions. Fig. 3(b) shows that the grayscale image is transformed by the affine matrix to a 350 pixels×200 pixels image stripped of useless environment information.

During the mixing process, it was likely that the SCC paste would splash on the inner wall of the mixer inlet and leave marks on it, which could be mistaken as a feature. Attention then was then paid to the ROI, which can truly provide useful features for the model to recognize shown in Fig. 3(c). The part of the image that is framed in a box was extracted as the ROI and resized to a certain size, 150 pixels×50 pixels.

Different illumination would cause differences in brightness between images, reflected in the grayscale values. If the grey level of an image were distributed unevenly, especially in a narrow range, the image contrast would be rather low, influencing the image sharpness and recognition. Thus, some part of the image would be either

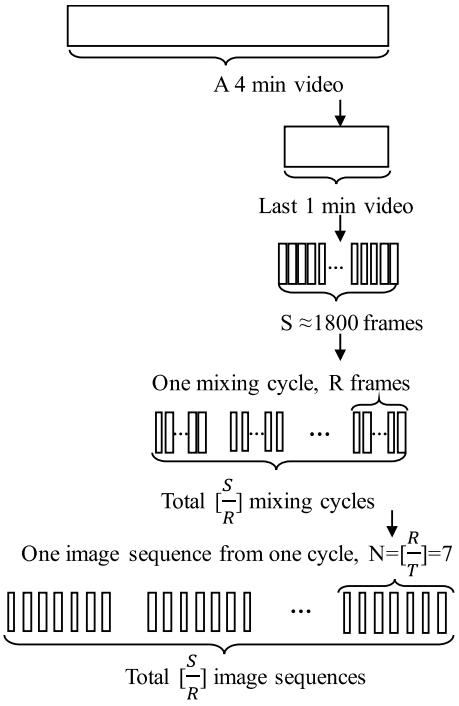


Fig. 5 The flowchart to enlarge dataset

too bright or too dark to distinguish. Consequently, the illumination will be a noise that should not be learned as a feature. Thus, histogram equalization was performed to address this problem, and the effect is shown in Fig. 3(d).

The four preprocessing steps were all applied in turn among these groups. Fig. 4 presents the preprocessing results of several images selected from several groups.

2.3 Data enlarging

The DL model needs a large amount of data, while the available data is restricted, so some processing steps were used to enlarge the amount of data. The steps to enlarge the data sets are shown in Fig. 5, from one video to multiple image sequences. First, only in the last one minute of the 4-minute mixing process was the mixture considered to be fully mixed and homogeneous (Cazacliu and Roquet 2009). Therefore, the last 60 s of the video was extracted and converted into many images in sequence for preprocessing. The number of these images was designated parameter S. For each video, the value of S was approximately 1800 frames.

After data preprocessing, it was easy to find that the images in sequence comprised some mixing cycles. A mixing cycle means the process in which the blades in a mixer rotate back to the initial position in the shortest time. The number of images in a mixing cycle was designated parameter R. For different mixers, R differed, as shown in Table 6. For the 60 L mixer, a mixing cycle contained 42 frames, implying that every time the mixer rotated one cycle, 42 frame images were generated. Then, the images in the sequence were divided by the mixing cycles

$$M = \frac{total\ frames}{frames\ in\ one\ mixing\ cycle} = \frac{S}{R} \tag{1}$$

Table 6 Down-sampling parameters of three groups

Group name	Images in one cycle (R)	Temporal resolution (T)	Image sequence length (N)
Group A	35	5	7
Group B, C, D	42	6	7

Table 7 The amount of data in the three groups

Groups	Number of samples	Number of video
A	10710	31
B	6468	21
C	10836	36
D	1848	6

M represents the number of mixing cycles in a 60 s video. Subsequently, the long image sequence was cut into M mixing cycles.

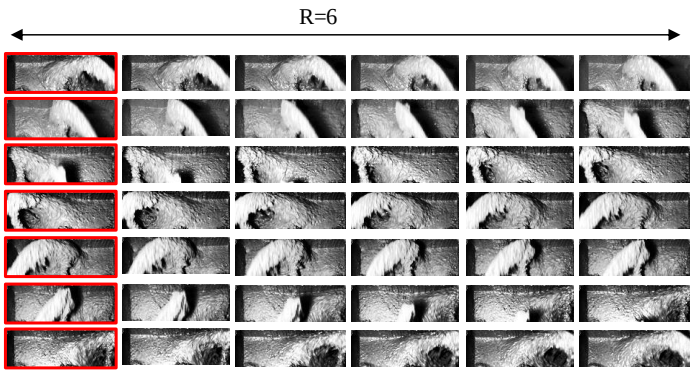
When all the images in the mixing cycles were put in the DL model, the computational cost was rather large to compute, and not all the frames were necessary because the differences between several adjacent images were fairly small, providing almost the same information. Therefore, some images carrying similar information in a cycle could be skipped. Thus, the parameter T, defined as the temporal resolution, was used to realize the downsampling operation by choosing every other T frame in a mixing cycle. Note that a larger T corresponds to more images being skipped and more time information being lost. Considering the computational cost and the time sequence information, the value of T should be neither too large nor too small. According to previous research (Ding and An 2018), a suitable value is between 5 and 9. The temporal resolution for each group in this paper is shown in Table 6. The number of frames extracted from one mixing cycle to form an image sequence was then converted from R into

$$N = \frac{the\ frames\ in\ a\ mixing\ cycle}{the\ time\ resolution} = \frac{R}{T} \tag{2}$$

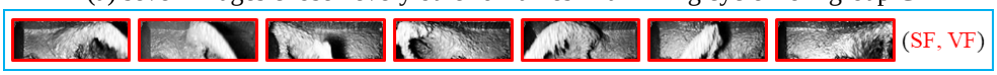
For each mixing cycle, the value of parameter N is 7, as shown in Table 6. The 7 chosen frames form one image sequence, namely, one sample. This process is shown in Fig. 6, taking one mixing cycle in group C as an example. The images framed in the box located on the first column are chosen every other six frames in a mixing cycle in group C as shown in Fig. 6(a) and eventually form one image sequence with seven images whose label contains the values of SF and VF, as shown in Fig. 6(b).

After enlargement, each video was first converted into a large number of frames. Then, these frames were turned into M mixing cycles, and each mixing cycle was transformed into one image sequence. Finally, each video was converted to M image sequences, corresponding to M samples. Each sample includes 7 preprocessed images with the SF and VF as its corresponding label vector.

Data augmentation was performed to further decrease the overfitting effect and enlarge the data. This method is clarified as follows. For each sample, the first image was placed at the end of the sequence in turn to maintain the temporal order. This approach works because the 7 images form a cycle, and whichever to be the starting is proper. Fig.



(a) seven images chosen every other 6 frames in a mixing cycle from group C



(b) one image sequence with SF and VF labels from group C

Fig. 6 Transition from a mixing cycle to an image sequence

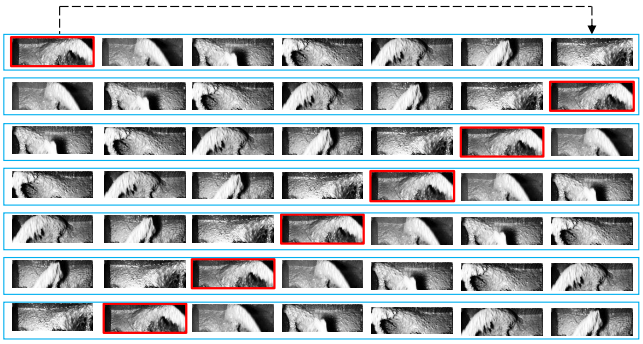


Fig. 7 Seven image sequences after operating data augmentation for group C

7 shows the augmentation operation. Six more image sequences derive from the original sequence, expanding the amount of initial data by a factor of 7.

In all, after data enlargement and augmentation, the amount of the data increases rapidly, as shown in Table 7, substantially mitigating the overfitting problem.

3. Dataset classification and model

After the data was collected, processing and enlarged, the samples should be divided into different data sets according to different goals. Then the training model was built in this part to train these processed data samples.

3.1 Dataset classification

The data were first classified into the training and validation set and the testing set. The former set was used to train the model and after training several times, the accuracy of the model improved considerably. Then the model was used to predict the testing set, and the predictions were compared with the actual labels, namely, the SF and VF values.

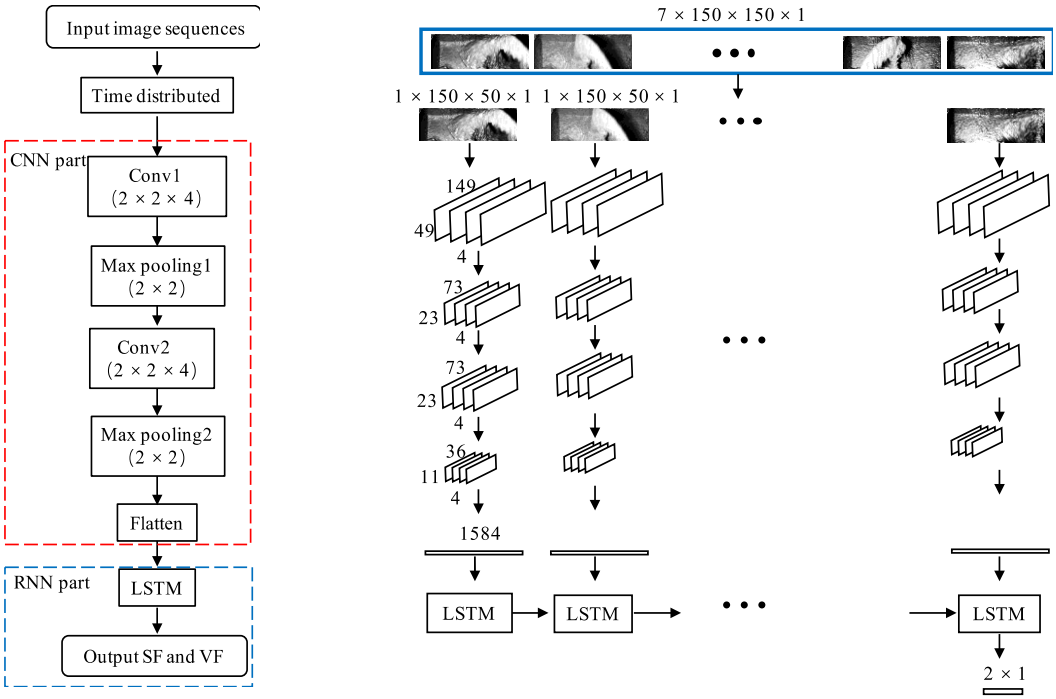
There were two main objectives in this study. For different goals, the dataset classifications were different.

For the first goal to estimate the workability of the SCC in different mixing conditions, the testing data should contain different mixers and mixing volumes, and the testing videos were randomly chosen from groups A, B and C, for a total of 26 groups as shown in Table 8. The left unchosen data of groups A, B and C automatically became the training and validation set. The mixing condition of the testing set was the same as that of one group in the training and validation set but with different workability.

The second goal is to demonstrate that a number of typical mixing conditions combined together can be used to

Table 8 The testing data from three groups

Testing set	Video number	SF (mm)	VF (s)
Test A	19	360	200
	20	485	69
	16	585	9
	1	600	13
	2	610	10
	3	610	18
	6	625	26
	4	648	13
	29	679	31
Test B	11	535	101
	13	555	66
	14	565	56
	1	600	8
	6	625	33
	7	645	35
	3	675	28
	4	700	10
Test C	15	350	12
	14	425	5
	20	495	35
	21	500	44
	16	548	9
	32	650	63
	7	685	9
	4	705	17
	31	708	67



(a) Model layers (b) Input and output of each layer

Fig. 8 Model architecture diagram

predict a new mixing condition. Here, the training and validation set was kept unchanged, formed by groups A, B and C. However, the testing data were a new set whose mixing condition was never seen or learned. Group D was the extra collected data to confirm this idea and was set as the testing data. It is emphasized that the mixing condition of the testing image sequences was different from that of the training and validation set.

If different kinds of mixers mixing different SCC volumes can be combined together to learn and achieve the abovementioned two goals, then whether or not the mixing condition of testing image sequences is included in the combined training set and learned by the model, this method can be applied to predict the workability of SCC in various mixing conditions with a combined training set.

3.2 Model architecture

As shown in Fig. 8, the model contains two main parts: the convolutional and the recurrent parts. The CNN part was used to extract the spatial features, and the recurrent part was used to extract temporal features. The two parts were connected by the time distributed layer because the recurrent part included an extra time dimension. Thus, the convolution parts could be applied for each of the time steps. CNN part comprised convolutional, pooling and flatten layers. The convolutional layers convolved with the input images or feature map with several local kernel filters. In this study, the convolutional layers both had four kernel filters with 2×2 kernel size, and the stride length was one, with no padding around the input data. The activation function was Rectified Linear Unit (ReLU). The max pooling layers were used to extract the most important features, with 2×2 size and stride length was one. Pooling reduced size of feature maps by about half. After convolutional and pooling, the flattened layer was applied to transform the high-dimensional feature map into a one-dimensional vector by thewidth or length of the feature map. The image sequences with 7 images contained not only spatial information extracted by CNN but also SCC time features for they were extracted in a whole mixing cycle in turn. To learn the sequential image features, the recurrent part was necessary and set as LSTM whose dimension was five to learn the temporal information from the sequences.

After the abovementioned layers were applied to the image sequences, a fully connected layer was finally used to transform the feature maps into a two-dimensional vector, comprising SF and VF, to achieve the regression goal. Cross entropy was employed as the cost function to measure the similarity between the predicted values and target values.

This model was implemented in Python using the Keras package. The CPU used was Intel(R) Core(TM) i7-7700 CPU, RAM 8.0 GB, and GPU was NVIDA GeForce GTX 1050.

4. Results and discussion

4.1 Prediction results in different mixing conditions

After the training process, the testing data were predicted. For one specific video, the predictions of the image sequences were computed together to obtain their average SF and VF values.

Table 9 Prediction results of the three learned groups

Testing set	Video number	Ground truth		Prediction		Relative error	
		SF (mm)	VF (s)	SF (mm)	VF (s)	SF (%)	VF (%)
Test A	19	360	200	419	161	16.4	19.5
	20	485	69	501	30	3.3	56.5
	16	585	9	589	17	0.7	88.9
	1	600	13	659	13	9.8	0
	2	610	10	574	21	5.9	110
	3	610	18	639	39	4.8	116.7
	6	625	26	628	43	0.5	65.4
	4	648	13	670	8	3.4	38.5
	29	679	31	647	43	4.7	38.7
Test B	11	535	101	507	65	5.2	35.6
	13	555	66	569	57	2.5	13.6
	14	565	56	524	50	7.3	10.7
	1	600	8	643	18	7.2	125
	6	625	33	633	22	1.3	33.3
	7	645	35	582	36	9.8	2.9
	3	675	28	625	21	7.4	25
	4	700	10	606	12	13.4	20
Test C	15	350	12	451	8	28.9	33.3
	14	425	5	417	8	1.9	60
	20	495	35	536	54	8.3	54.3
	21	500	44	460	47	8	6.8
	16	548	9	561	17	2.4	88.9
	32	650	63	676	56	4	11.1
	7	685	9	670	36	2.2	300
	4	705	17	635	15	9.9	11.8
	31	708	67	688	51	2.8	23.9

Table 9 shows the ground truth, predicted SF values, estimated VF values and relative error of the testing set in the three groups. It can be concluded that the training set combined by three mixing conditions chosen from groups A, B and C can effectively predict the value of SF. The prediction relative errors of the VF are numerically larger because the value of the ground truth is relatively small, the same magnitude as the absolute error; therefore, the results of the VF were influenced. For the same fluctuations in number, VF was influenced more than SF.

However, whether the workability of SCC was qualified is more important in practice. The qualified SF value should be over 600 mm. And for the VF, the value should be between 5 s and 25 s.

If the SF or VF of the ground truth is qualified but that of the prediction is not, then the prediction is considered inaccurate, as marked by an underline in Table 9, vice versa. Only if the SF and VF are qualified or unqualified both in ground truth and prediction is the prediction considered accurate.

The relation between ground truth and prediction is shown in Fig. 9. The results also show whether the SF and VF are predicted accurately. Taking the SF prediction as an example, 600 mm is a boundary to distinguish the qualified and unqualified SCC. The dots in the lower left and the upper right dividing by the line 600 mm in Fig. 9(a) are estimated accurately, while the dots in the lower right and the upper left is predicted inaccurately. So is the VF and the dividing boundary is 25 s. For SF prediction of test A, three dots are distributed in the lower left and four in the upper right. They are all estimated accurately. Only one dot in the lower right is predicted inaccurately.

Therefore, the accuracy of the SF and VF in test A is both 89%, shown in Fig. 9(a)-(b). The corresponding results for test B are 85% and 75%, presented in Fig. 9(c)-(d) and those of test C are 100% and 89% as shown in Fig. 9(e)-(f). The total accuracy of the 26 dots is shown in Fig. 9(g-h). The SF and VF accuracy assessments are 92% and 85%, respectively. The total 26 groups of testing data show that the model can accurately estimate the workability of SCC in different mixers and mixing volumes, constituting different mixing conditions.

It can be concluded that the combined training set can effectively predict SCC workability. This result shows that the model can estimate several kinds of testing data at the same time whose mixing condition is the same as that of one group in the training set. The mixer of group A is different from that of group B and C, so it is easy to distinguish group A from group B and C. As to group B and C, although their mixers are the same, their mixing volume differs a lot which makes it easier for the model to differentiate.

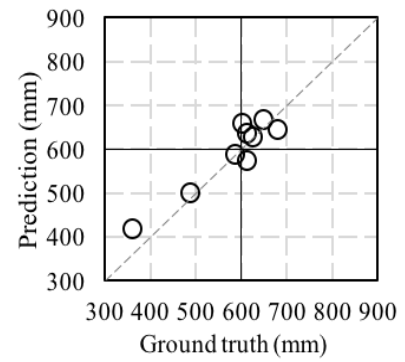
Table 10 Prediction results of the unlearned group D with a new mixing condition

Testing set	Ground truth		Prediction		Relative error	
	SF (mm)	VF (s)	SF (mm)	VF (s)	SF (%)	VF (%)
Test D	660	22.5	670	30.1	1.6	33.9
	550	30.7	523	38.8	4.9	26.3
	455	21.2	593	21.3	30.3	0.4
	465	17.6	599	25.0	28.7	42.2
	655	68.4	612	50.0	6.6	26.8
	660	55.2	600	26.4	9.1	52.3

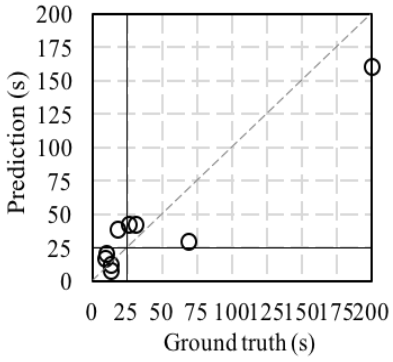
4.2 Prediction result of the unlearned group with a new mixing condition

Group D is the unlearned group, whose mixing condition is not contained in the training set. In addition, six videos of group D were collected in this paper sets as the extra testing set, with mixing conditions different from those of groups A, B, and C which means its mixing condition differed from the combined training set. Table 10 shows the ground truth, prediction and relative error of group D. Although the VF value of the first group was predicted to be unqualified, as marked by an underline, the left ones are all predicted correctly, consistent with the ground truth.

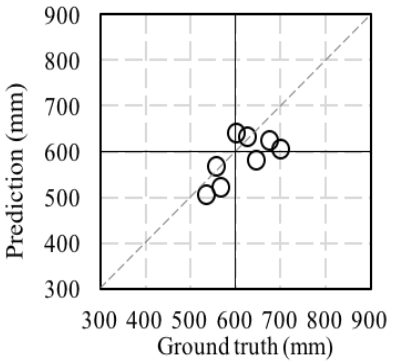
Fig. 10 shows the relation between the ground truth and prediction as well as the prediction accuracy. The SF and VF prediction accuracies of group D, whose mixing conditions are different from those of the training sequences, are 100% and 83%, respectively. Although the accuracy of the VF is relatively low, as shown in Fig. 10, the distance between the ground truth and prediction is smaller. The mixer of group C is the same as that of group



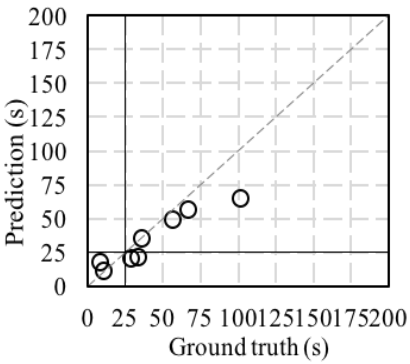
(a) SF prediction of test A



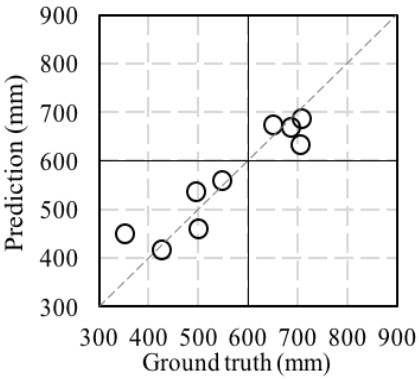
(b) VF prediction of test A



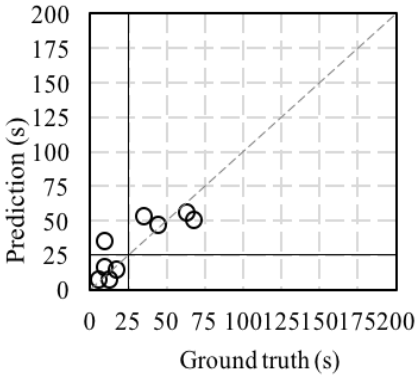
(c) SF prediction of test B



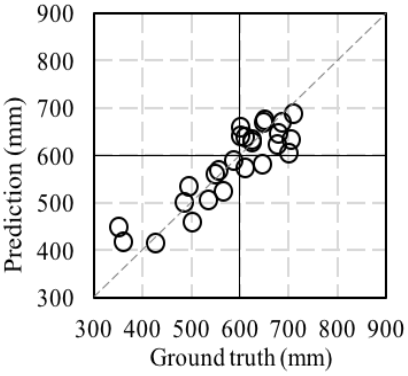
(d) VF prediction of test B



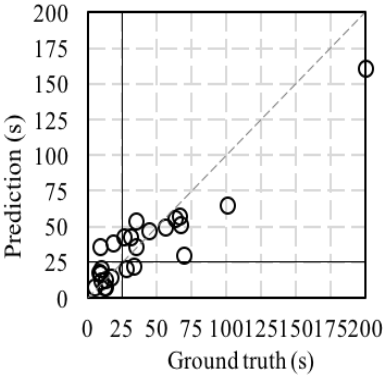
(e) SF prediction of test C



(f) VF prediction of test C



(g) All the SF prediction results

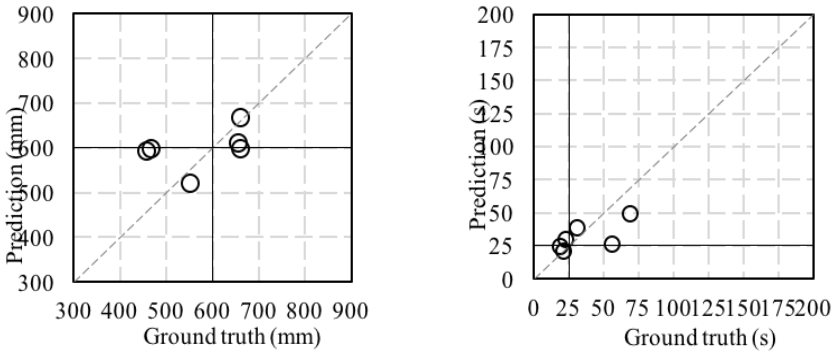


(h) All the VF prediction results

Fig. 9 SF and VF prediction results in the learned groups

B and D, and the mixing volume is between them. The feature of group C is similar to group B and D. Therefore, although there are not any samples of group D in the training set, the model still has the ability to make a good prediction on group C.

Therefore, it can be concluded that the combined training and validation set in the DL model can estimate the unlearned group D well.



(a) SF prediction of test D (b) VF prediction of test D

Fig. 10 SF and VF prediction results in the unlearned group

5. Conclusions

In this paper, a method was proposed to estimate the workability of SCC with different types of mixers and mixing different volumes by recording the mixing process, with four groups in total and approximately thirty thousand samples. The SCC mixing videos were transformed into a series of image sequences to fit the DL model to predict the SF and VF values of SCC. The model achieved good prediction accuracy of the testing data.

- It can estimate the testing set with different kinds of mixing conditions at the same time when their mixing conditions, namely, the mixer and mixing volume, have been learned and contained in the training set.
- In addition, the model also achieved the goal of estimating image sequences whose mixing condition was not learned and included in the training set. Therefore, the training set combined by three different mixing condition groups can be used to estimate SCC with different workability at the same time regardless of whether the mixing condition is included in the training set and learned by the model.

New mixing conditions can be predicted through the use of typical mixing conditions. Therefore, the work to collect the mixing conditions is reduced and proven to be feasible. The model can be extended in the future to more kinds of mixers mixing different volumes. The DL model can be used in other kinds of mixers, such as a twin-shaft mixer and vertical paddle mixer, and with different mixing volumes. In the next step, The plan is to collect further video data to expand the use of this technique and establish an automatic prediction system to estimate the workability of SCC in various mixing conditions.

Acknowledgments

This work was supported by Open Research Fund Program of State key Laboratory of Hydroscience and Engineering (sklhse-2019-C-05).

References

Abd, A.M. and Abd, S.M. (2017), "Modelling the strength of lightweight foamed concrete using support vector machine (SVM)", *Case Stud. Constr. Mater.*, **6**, 8-15. http://dx.doi.org/10.1016/j.cscm.2016.11.002.

Affonso, C., Debiaso Rossi, A.L., Antunes Vieira, F.H. and de Leon Ferreira de Carvalho, A.C.P. (2017), "Deep learning for biological image classification", *Exp. Syst. Appl.*, **85**, 114-122. http://dx.doi.org/10.1016/j.eswa.2017.05.039.

An, X., Wu, Q., Jin, F., Huang, M., Zhou, H., Chen, C. and Liu, C. (2014), "Rock-filled concrete, the new norm of SCC in hydraulic engineering in China", *Cement Concrete Compos.*, **54**, 89-99. http://dx.doi.org/10.1016/j.cemconcomp.2014.08.001.

Beaupré, D. (1994), "Rheology of high performance shotcrete", Ph. D. Thesis, Uni. Of British Columbia, Canada.

Bengio, Y., Courville, A. and Vincent, P. (2013), "Representation learning: A review and new perspectives", *IEEE Tran. Pattern Anal. Mach. Intel.*, **35**(8), 1798-1828. http://dx.doi.org/10.1109/tpami.2013.50.

Cabaret, F., Bonnot, S., Fradette, L. and Tanguy, P.A. (2007), "Mixing time analysis using colorimetric methods and image processing", *Indus. Eng. Chem. Res.*, **46**(14), 5032-5042. http://dx.doi.org/10.1021/ie0613265.

Cazacliu, B. and Roquet, N. (2009), "Concrete mixing kinetics by means of power measurement", *Cement Concrete Res.*, **39**(3), 182-194. http://dx.doi.org/10.1016/j.cemconres.2008.12.005.

Chopin, D., Cazacliu, B., de Larrard, F. and Schell, R. (2007), "Monitoring of concrete homogenisation with the power consumption curve", *Mater. Struct.*, **40**(9), 897-907. http://dx.doi.org/10.1617/s11527-006-9187-8.

Chou, J.S., Tsai, C.F., Anh-Duc, P. and Lu, Y.H. (2014), "Machine learning in concrete strength simulations: Multi-nation data analytics", *Constr. Build. Mater.*, **73**, 771-780. http://dx.doi.org/10.1016/j.conbuildmat.2014.09.054.

Daumann, B. and Nirschl, H. (2008), "Assessment of the mixing efficiency of solid mixtures by means of image analysis", *Powd. Technol.*, **182**(3), 415-423. http://dx.doi.org/10.1016/j.powtec.2007.07.006.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L. (2009), "Imagenet: A large-scale hierarchical image database", *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255.

Ding, Z. and An, X. (2017), "A method for real-time moisture estimation based on self-compacting concrete workability detected during the mixing process", *Constr. Build. Mater.*, **139**, 123-131. http://dx.doi.org/10.1016/j.conbuildmat.2017.02.047.

Ding, Z. and An, X. (2018), "Deep learning approach for estimating workability of self-compacting concrete from mixing image sequences", *Adv. Mater. Sci. Eng.*, **2018**, Article ID 6387930, 16. http://dx.doi.org/10.1155/2018/6387930.

Donahue, J., Hendricks, L.A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K. and Darrell, T. (2017), "Long-term

recurrent convolutional networks for visual recognition and description", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* , 2625-2634.

Gidaris, S. and Komodakis, N. (2015), "Object detection via a multi-region and semantic segmentation-aware cnn model", *Proceedings of the IEEE International Conference on Computer Vision*, 1134-1142.

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014), "Rich feature hierarchies for accurate object detection and semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 580-587.

Hariri-Ardebili, M.A. and Pourkamali-Anaraki, F. (2018), "Support vector machine based reliability analysis of concrete dams", *Soil Dyn. Earthq. Eng.*, **104**, 276-295. http://dx.doi.org/10.1016/j.soildyn.2017.09.016.

He, K., Zhang, X., Ren, S. and Sun, J. (2014), "Spatial pyramid pooling in deep convolutional networks for visual recognition", *IEEE Tran. Pattern Anal. Mach. Intel.*, **37**(9), 1904-1916..

Hutchinson, T.C. and Chen, Z.Q. (2006), "Improved image analysis for evaluating concrete damage", *J. Comput. Civil Eng.*, **20**(3), 210-216. http://dx.doi.org/10.1061/(asce)0887-3801(2006)20:3(210).

Lee, B.Y., Kim, Y.Y., Yi, S.T. and Kim, J.K. (2013), "Automated image processing technique for detecting and analysing concrete surface cracks", *Struct. Infrastr. Eng.*, **9**(6), 567-577. http://dx.doi.org/10.1080/15732479.2011.593891.

Li, S. and An, X. (2014), "Method for estimating workability of self-compacting concrete using mixing process images", *Comput. Concrete*, **13**(6), 781-798. http://dx.doi.org/10.12989/cac.2014.13.6.781.

Marinoni, N., Pavese, A., Foi, M. and Trombino, L. (2005), "Characterisation of mortar morphology in thin sections by digital image processing", *Cement Concrete Res.*, **35**(8), 1613-1619. http://dx.doi.org/10.1016/j.cemconres.2004.09.015.

Okamura, H. and Ouchi, M. (2003), "Self-compacting concrete", *J. Adv. Concrete Technol.*, **1**(1), 5-15. http://dx.doi.org/10.3151/jact.1.5.

Taffese, W.Z. and Sistonen, E. (2017), "Machine learning for durability and service-life assessment of reinforced concrete structures: Recent advances and future directions", *Auto. Constr.*, **77**, 1-14. http://dx.doi.org/10.1016/j.autcon.2017.01.016.

Taffese, W.Z., Sistonen, E. and Puttonen, J. (2015), "CaPrM: Carbonation prediction model for reinforced concrete using machine learning methods", *Constr. Build. Mater.*, **100**, 70-82. http://dx.doi.org/10.1016/j.conbuildmat.2015.09.058.

Wu, Q. and An, X. (2014), "Development of a mix design method for SCC based on the rheological characteristics of paste", *Constr. Build. Mater.*, **53**. http://dx.doi.org/10.1016/j.conbuildmat.2013.12.008.

Yaseen, Z.M., Deo, R.C., Hilal, A., Abd, A.M., Bueno, L.C., Salcedo-Sanz, S. and Nehdi, M.L. (2018), "Predicting compressive strength of lightweight foamed concrete using extreme learning machine model", *Adv. Eng. Softw.*, **115**, 112-125. http://dx.doi.org/10.1016/j.advengsoft.2017.09.004.

Yokoyama, S. and Matsumoto, T. (2017), "Development of an automatic detector of cracks in concrete using machine learning", *Procedia Eng.*, **171**, 1250-1255.

Zhang, A., Wang, K.C. P., Li, B., Yang, E., Dai, X., Peng, Y., Fei, Y., Liu, Y., Li, J.Q. and Chen, C. (2017), "Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network", *Comput. Aid. Civil Infrastr. Eng.*, **32**(10), 805-819. http://dx.doi.org/10.1111/mice.12297.