Author response to reviews of: Paper 17 - Real-time analogue gauge transcription on mobile phone

We would like to first thank the reviewers for their suggestions and kind consideration of this paper. Here we aim to address points raised by both the reviewers and meta-reviewers.

Reviewer 1

Possible insufficient prior work. The number of prior works is limited. Probably due to a lack of data for benchmarking this task (something which we hope our paper addresses). We believe our related work covers the vast majority of prior work.

Clarify details of the error metrics. We have added extra details on the evaluation metrics in the text and also provide these details below:

- *Detection Accuracy (Det. Acc.)* is defined as the proportion of images in which a gauge is correctly detected, for the Real-Gauges dataset a gauge is present in every image.
- *The read return rate* $R_{\%}$ is used to measure the proportion of frames where the system returns a reading (regardless of value).
- *Absolute read error* is measured as the mean μ_V and the standard deviation σ_V of the difference between the ground truth gauge reading and the system prediction.
- *Absolute relative read error* is measured as the mean μ_R and the standard deviation σ_R, and represents absolute read error as a proportion of the gauges reading range, to demonstrate how significant the error is in terms of each gauge.
- *Absolute pointer angular error* is measured as the mean, μ_θ, and standard deviation, σ_θ, which measure the difference between ground truth angle of the pointer and the system prediction.

Table 1 is not referenced. Thank you for pointing this out. We have now included a short discussion on Table 1 in the results section.

Effect of gauge read error with respect to perspective correction error. There is a correlation between read error and errors in perspective correction. In the paper we analyse this and observe that increasing the angle between the camera and gauge face increases both read and projection error (see Figure 8 and 9). Figure 9, shows that error in perspective correction is low enough for good read performance if the absolute angle is less than 20 degrees.

Failure case discussion. We do make discussion to reasons for failure in both the gauge detection and gauge reading experiments. However, we have also now added further failure case discussion for the perspective recovery experiment.

Reviewer 2

Explanations are too succinct particularly for details on perspective correction. We have extended the explanation of perspective correction in the text (section 2, stage 2).

Add links to the public image repositories holding the new datasets. We are still in the process of setting up a server for holding the datasets. This will be online in time for the workshop event.

Memory requirements were not evaluated in regards to the decision use a low res image rather than high res. We have added a brief discussion to the paper that shows that by using a lower resolution image we were able to increase model speed by a factor of 5.5, this was vital for running the model in real time on mobile.

Was solely synthetic data used during training? Yes we do only use synthetic data for training (no real images are seen). This is made reference to in Section 3 - Training. However, we have now further highlighted this in the introduction.

Recommendation of including an experiment in which the gauge pointer varies between the minimum and maximum measurement range by quantifying the theta angle measurement error. We believe that this experiment was conducted in the original paper, in the Gauge reading experiment. The relationship between angle error and absolute value error is now further clarified in the text to help avoid confusion. We would also like to note that for some gauges reaching maximum value for the gauge pointer was not feasible.

Baseline not competitive. We argue the baseline method is competitive and certainly performs well on some meters in the Kaggle dataset. It is also one of the most recent and open source methods. We note that many prior methods perform well on datasets they were trained on but do not generalise well to other data. In our work we aim to address this using synthetic training data, showing our method generalises across different datasets. We clarify this point in the text.

Translate the reading into a view of the full scale. This is a nice suggestion and we will look into incorporating it into our system.

Meta-Reviewers

Frameworks used. The model was built and trained using the PyTorch framework. For iOS deployment we converted to a 16bit CoreML model. The model is capable of running on the Neural Engine. We tested on the iPhone 11 using the Neural Engine and achieved an average speed of 25fps. Running using only the GPU on the iPhone 7, still achieves very good performance of around 20fps. These points have also been added to the text.