NRC-GAMMA: A Large Novel Open-Access Gas Meter Image Dataset

Ashkan Ebadi National Research Council Canada
Patrick Paul National Research Council Canada
Sofia Auer National Research Council Canada
Stéphane Tremblay National Research Council Canada
Email: {ashkan.ebadi, patrick.paul, sofia.auer, stephane.tremblay}@nrc-cnrc.gc.ca

Abstract

Automatic meter reading technology is not yet widespread. Gas, electricity, or water accumulation meters reading is mostly done manually on-site either by an operator or by the homeowner. With the recent advances in the fields of artificial intelligence and computer vision, automatic meter reading systems are becoming more viable than ever. In this work, we introduce a novel large benchmark open-access dataset of real-life gas meter images, called NRC-GAMMA. We employed a systematic approach to label the images, validate the labellings, and assure the quality of the annotations. The dataset contains 28,883 images of the entire gas meter along with 57,766 cropped images of the left and the right dial displays. We hope the NRC-GAMMA dataset helps the research community to design and implement accurate, innovative, intelligent, and reproducible automatic gas meter reading solutions.

1 Introduction

Natural gas is an economical energy source that is also relatively clean-burning and environment friendly [1]. For decades, the gas has been supplied to residential and industrial places, being used in many different ways including as a cooking and heating fuel in households, powering industrial furnaces, or even fueling vehicles [2]. In Canada, although the gas consumption differs across the country, the main residential use of natural gas is for space and water heating purposes [3]. In 2018, Canada consumed an average of 11.2 billion cubic feet per day of natural gas, having Alberta and Ontario provinces as the largest gas consumers [4].

Gas meters are installed to monitor the consumed gas quantity to charge the consumers accordingly. In most countries, the gas meter reading is a manual process being performed by an operator on-site [5], visually checking the meter and entering the consumption amount manually [6], and taking a proof image of the meter for the record [7]. The proof image may be later used to verify the reading offline with another operator [7]. Such a method is highly inefficient, time-consuming, expensive, and prone to errors [6, 8]. Intelligent automated real-time gas meter reading systems are highly needed to overcome the limitations of the manual process. Such real-time reading systems can be beneficial for both the consumers and the provider companies. From a consumer perspective, it would help them, for instance, to reduce or better manage the consumption and the cost. For a gas provider company, the system would assist them to better understand usage profiles, leaks, and to improve usage predictions.

Although smart meter reading devices are one of the alternative solutions for automatically recording gas (or other utilities) consumption, in many countries, they are not yet widely available and the reading is still performed manually [7]. In addition, despite clear advantages, smart meters are very expensive. For example, the cost of transition from non-smart meters to smart ones in Australia was estimated to be about $ 1.6 billion with a cost burden carried to the customers to pay for the upgrade [9]. Therefore, many countries have delayed the smart meter transition program. With recent advances in artificial intelligence (AI) and computer vision, alternative digital solutions are now more feasible than ever. For example, an automatic meter reading (AMR) system can be designed by installing an image acquisition device, e.g., a camera, in the meter box, and training a machine/deep learning model on the captured real-time images of the meter readings [10]. Image-based AMRs are cost-efficient, as it is not required to change the meter, and they can be installed quickly [11]. They are also scalable and can generalize to different applications and styles of the meters [10].

Although several studies in the literature proposed/developed either image-based AMR systems or learning methods for a specific task in the system (e.g., text detection in the whole image of the meter or digit segmentation), most of the AMR-related datasets used to build the models are not publicly available as images of the meters often belong to a service company [8]. The few existing public meter datasets mostly contain images of digit-based meters, also called counter meters or cyclometers (e.g., [8, 12]. This limitation would jeopardize the reproducibility of the performed experiments as well as the possibility of a thorough assessment of proposed systems [13]. Additionally, collecting properly and annotating a large amount of data required for training (deep) learning models is an expensive task restricting research individuals/teams with limited financial resources.

To overcome these limitations and inspired by the open-source activities of the scientific community, we introduce a novel large benchmark image dataset of a residential diaphragm gas meter, called NRC-GAMMA, that is available to the public at the NRC-GitHub. The NRC-GAMMA dataset contains 28,883 images of the entire gas meter along with 57,766 cropped images of the left and the right dial displays, carefully annotated and systematically validated. To the best of our knowledge, this is the largest open-access dataset of a residential diaphragm gas meter that employed well-defined annotation and validation protocols. We hope the dataset helps researchers and innovators in building deep learning models and/or AI-powered meter reading components/systems to tackle the problem of reading residential meters automatically and accurately with the ultimate goal of proposing intelligent and efficient AMR methods/systems.

2 NRC-GAMMA Dataset

The data was collected from an Itron 400A gas meter installed and used in one of the test house facilities of the National Research Council of Canada (NRC), built in 2009 and located in Ottawa, Ontario. The gas meter is a light commercial (also suitable for high-load residential) gas diaphragm meter with a capacity of $400ft^3/hr$. The meter has two proving dial displays rotating anticlockwise at different speeds, the left one at $50dm^3$ and the right one at $10dm^3$ per revolution, and a 5-digit cyclometer display. We intentionally chose this type of gas meter for at least three reasons: 1) By having continuous monitoring of the dial displays, better gas management for both the consumer and gas providers would be possible, 2) A well-annotated large-scale dial display meter dataset is not publicly available to the best of our knowledge, and 3) There are other types of utility in addition to the gas meters that only have dial displays. Having a labeled set of dial displays would expand the application of the NRC-GAMMA dataset to other utility meters as well.

To ensure that the images are captured continuously and consistently, we built a smart image capturing system powered by advanced edge processing techniques able to capture images in different conditions during day and night. The system contained a Raspberry Pi 3b+, i.e., a small single-board computer, powered by a Power over Ethernet (PoE) HAT that allows powering the Raspberry Pi board using PoE–enabled networks, a 5-megapixel infrared camera module (NoIR) that enabled us to take quality pictures at night as well using infrared lighting. The components were all boxed in a weatherproof openH Rubicon IP67 case. The images were taken automatically on average every ~ 3 seconds on the 20^{th} of January 2020 between 0:05 and 23:59 Eastern Time, with the average temperature of -14.8 ℃ and wind blowing with the average speed of $11km/hr$.

The NRC-GAMMA dataset contains 28,883 raw images of the entire gas meter along with cropping metadata and script to create 57,766 cropped images of proving dials from the original images, i.e., 28,883 images per dial. That is the dataset includes two types of images: 1) the raw images of the gas meter, and 2) cropped images of the left and right dials, separately. The cropped images (160×160 pixels) only show the dial region, and were included to facilitate researchers' analytic efforts in case they would like to only analyze the dial displays rather than the whole image of the gas meter. The dataset is publicly available at the NRC-GAMMA repository. In the GitHub repository, users are provided with a Python notebook

Fig. 1: a) A sample raw image of the gas meter panel captured at night, b) cropped image of the respective left dial, and c) cropped image of the respective right dial.

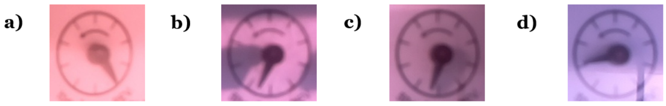

Fig. 2: a) A sample cropped image with no artifact, b) with heavy shadow, c) blurry due to windy condition, and d) glare due to direct sunbeam reflection.

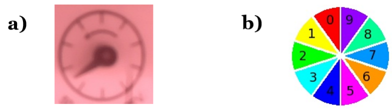

Fig. 3: A sample annotator's task that contained two parts, a) a sample dial image, and b) the interface in which the annotator should select the representative region based on the given dial image. In this example, the correct answer is 3.

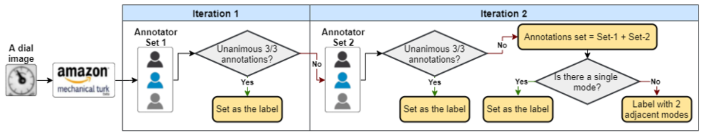

Fig. 4: The high-level conceptual flow of the decision logic for setting labels of the gas meter dial images.

that contains all the steps required to download the dataset archive file, extract the original gas meter images as well as the cropped ones, and store them on their local device. The provided scripts are well-documented allowing users to modify the script, e.g., use different cropping parameters, based on their research objectives and requirements if needed. Fig. 1 shows a sample gas meter image along with cropped regions of the left and right dials.

Since the images were taken during the whole day, the camera was affected by various environmental factors, e.g., varying lighting conditions and wind speed that resulted in various sets of artifacts in the captured images. We intentionally created this situation since these diverse variations in the dataset provide researchers with a real-life example allowing them to stand for variations and create flexible and reliable AMR systems that fit real-life conditions. Fig. 2 shows a sample image with no artifact plus some sample captured images with various observed artifacts such as shadow and light reflection.

3 The Annotation Process

To annotate and label images in the NRC-GAMMA dataset, we used Amazon Mechanical Turk (MTurk) which is a crowdsourcing service to perform discrete on-demand tasks. We created a user-friendly interface (Fig. 3) facilitating the annotation process for annotators while decreasing the margin of error by providing an easy-to-use graphical user interface (GUI). Given a cropped image of a dial display, the annotator was tasked to click on the respective region in the interface that in their opinion was the most representative annotation for the image. To avoid ambiguity and to follow a consistent process, we instructed annotators to take the next region (anticlockwise) as the label in cases where they were uncertain between two regions. We created random batches of 1,000 images of the dials, i.e., 58 batches in total. Each image was annotated by at least 3 annotators. A large proportion of the images were annotated by 6 annotators. In the rest of this section, we explain in detail the annotation process and the systematic approach that we followed to set the final labels for the images and to ensure the quality of the labels.

Our defined annotation process accounts for subjectivity and human error and reduces their impact to a minimum. Fig. 4 shows a high-level conceptual flow of the decision logic. The high-level annotation logic for a given dial image is based on two iterations and is described as follows:

1. In the first iteration, the given dial image is sent to three distinct annotators.
   a. When all three annotations have unanimous consensus, i.e., three annotations are identical, it is set as the label of the image and the process is terminated for the given image.
   b. Otherwise, the annotations are kept as Set-1 and we go to Iteration 2.
2. In the second iteration, the image is sent to three new distinct annotators.
   a. If all three annotations received from the new annotators are unanimous, the annotation is set as the label of the image, and the process is terminated.
   b. Otherwise,
      i. The new annotations are stored as Set-2.
      ii. Set-1 and Set-2 are merged and named as the annotations set.
      iii. If there is a unique majority mode (without ties) in the annotations set, then this mode is set as the label, and the process is terminated.
      iv. If the given image has a tie in the mode, whereby the two modes are adjacent, then the image is labelled with the 2 adjacent modes, and the process is terminated.
      v. All other images not labelled at this point, are visually inspected and labelled accordingly.

We systematically monitored and verified the quality of annotators and annotations during the labelling procedure. The quality control and data integrity checks were performed by coding scripts in R programming language. The R scripts are available upon request. The annotation difficulty level varies for various reasons, e.g., with the position of the hand in the dial display or the time of day, hence; evaluating annotators is not simple. We assessed the quality of annotators, after receiving the first iteration of annotations. First, we define an *unacceptable* annotation if one of the following two conditions occurs: 1) it is an annotation for an image where there is no consensus or majority in the annotation set, e.g., $\{4,8,3\}$, or 2) it is an annotation related to an image where there is a majority mode, and it is the annotation that did not make up that majority, and also is not adjacent to the majority mode value for that image. For instance, if the 3 annotations of an image are $\{1,1,3\}$, the 3 is considered unacceptable as it is not the majority mode and it is not adjacent to the majority mode, i.e., ± 1 of the majority mode of 1. Had the last annotation been a 0 or a 2 (instead of the 3) then that annotation would have been considered acceptable. Second, the total number of annotations completed by each annotator was calculated, and the quality of an annotator was measured using the following metric: $Q_a = \frac{A_u}{A_t}$, where Q_a is the quality of a given annotator, A_u is the total number of unacceptable annotations, and A_t is the total number of annotations completed by an annotator. We grouped the annotators based on the quality of their annotations (i.e., their Q_a) into three categories: 1) high quality, if $Q_a < 0.2$, 2) fair, if $0.2 \leq Q_a \leq 0.5$, and 3) low quality, if $Q_a > 0.5$. In other words, an annotator with low quality of work will assign more than half of its images with an unacceptable label. In those cases, we preferred to not set the label solely based on its annotations and considered another set/round of annotation to make the decision. Of the 57,766 images sent for the first round of annotations, 1,102 annotators produced 173,298 annotations, of which 10,026 annotations (6%) were considered unacceptable. Among those annotators, 10% were classified as low quality, and 90% were annotators with fair or high quality of work (5% and 85% respectively).

Fig. 5 shows the detailed annotation methodology with the corresponding number of images at each step. As seen, the process

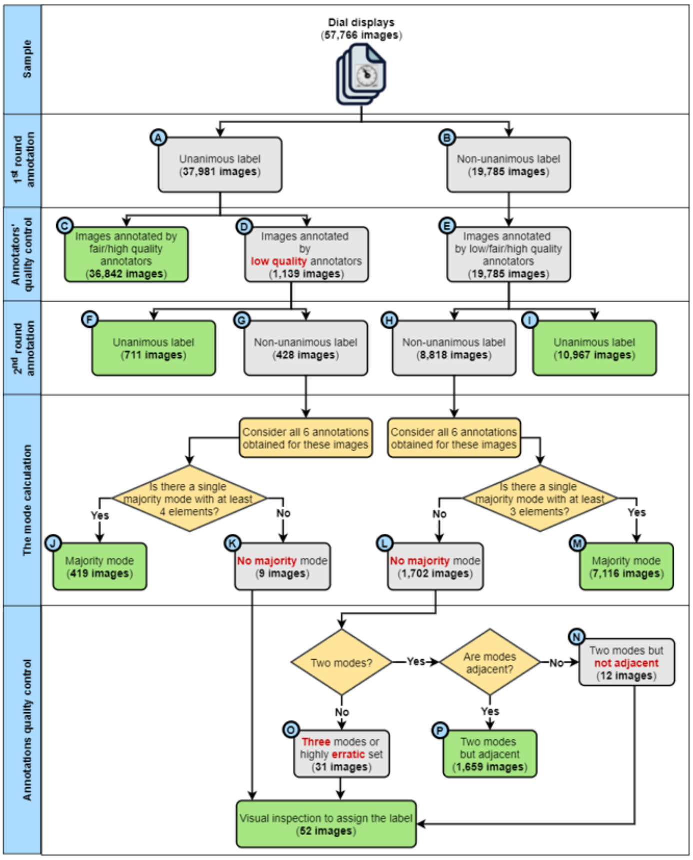

Fig. 5: Detailed annotation methodology with the corresponding number of images at each step.

started with 57,766 cropped images of the left and right dial displays (28,883 images per dial). In the first round of annotation, images were sent to three annotators in the Mechanical Turk platform and three annotations were received for each image, i.e., 173,298 annotations in total. From 57,766 images, 37,981 images (65.7%) were annotated with a unanimous label that is all three annotators agreed on the label. For the rest, i.e., 19,785 images, the annotations were not unanimous that is there was at least one disagreement. In the next step, we checked the quality of annotators using the quality metric defined in the previous section (Q_a). For the images with a unanimous label in the first round of annotation (n=37,981, left branch in Fig. 5, box A), 36,842 images (97%, box C in the figure) were annotated by annotators with fair/high quality ($Q_a \leq 0.5$). The unanimous label was set and the annotation process terminated for this set of images. For the remaining 1,139 images (3%, box D in the figure) with unanimous label quality of at least one annotator was not satisfactory, hence, the set remained in the process. Images with a non-unanimous label in the first round of annotation (n=19,785, right branch in Fig. 5, box B), were also considered for further verification and remained in the process, regardless of the quality of annotators (box E in Fig. 5), however, we restricted annotators with low/fair quality from annotating again.

We did a second round of annotation for the remaining images, i.e., all 20,924 images that did not have unanimous annotations (19,785, box E in Fig. 5) or those from low-quality annotators (1,139, box D). For images annotated by at least one annotator with low quality in the first round of annotation (box D), if we obtained a unanimous label in the second round, it was set as the label and the annotation process terminated for this set of images (n=711, box F in Fig. 5). We kept the remaining images (n=428, box G) in the process for further verification. We followed the same process for images with a non-unanimous label in the first round of annotation (box E). Out of these 19,785 images, a unanimous label obtained in the second round of annotation for 10,967 images (55%), it was set as the label, and the annotation process terminated for this set of images (box I in the figure). The remaining images (n=8,818, box H) were kept in the process. Up to this step, a label was set for 48,520 images (84%). We followed a careful quantitative/manual process for the remaining 9,246 images (16%) to set the label. For these images (boxes G and H in Fig. 5), we first combined the annotations received in both rounds, i.e., $3+3$ annotations per image. For images with non-unanimous annotation in the second round and unanimous annotation but at least with one low-quality annotator in the first round of annotation (box G in Fig. 5), we checked if there exists a single, unique, no tie, majority mode with at least 4 elements in the set of 6 annotations. If such mode was found, it was assigned as the label for those images, and the process terminated (n=419, box J in Fig. 5). For images with a non-unanimous annotation in both rounds of annotation (box H in Fig. 5), we checked if there exists a single, unique, no tie, majority mode with at least 3 elements in the set of 6 annotations. If such mode was found, it was assigned as the label for those images, and the process terminated (n=7,116, box M in Fig. 5). We applied a stricter rule for images in box G compared with those in box H (i.e., mode with at least 4 elements vs. mode with at least 3 elements) as we identified some low-quality annotators in the first round of annotation for images in box G despite agreeing on a unanimous label. After this step, a label was set for 56,055 images (97%).

We did further verification for the remaining 1,711 images (3%). Those 9 images in the left branch of the decision tree with no majority mode (box K in Fig. 5) were sent for visual inspection. For 1,702 images in box L in the figure, we first checked each image if there exist two modes in the set of 6 annotations. If yes, we checked if those two modes refer to adjacent zones in the dial display, and if they were adjacent, we duplicated those images and assigned one of the modes to each of the duplicated images, and terminated the process for them (n=1,659, box P in Fig. 5). If the two modes were not adjacent, we sent those images for visual inspection (n=12, box N in Fig. 5). The reason for the adjacency condition was to ensure that the main cause of disagreement between annotators is the difficulty level of the image and not a careless/random annotation. Also, if there were more than two modes, or there is one mode but the annotation set is highly erratic that is there are so many other values (e.g., 2, 2, 1, 3, 5, 9), or no mode at all in the set of 6 annotations (n=31, box O in Fig. 5), the image was sent for visual inspection. Overall, 52 images were sent for visual inspection (from boxes K, N, O in Fig. 5). In the visual inspection step, images were annotated by three subject experts from our team, and the majority voting rule was applied to assign the final label.

4 Conclusion

Smart gas meters enable automatic reading of gas consumption, however, they are not yet being widely deployed even in the developed countries, having many conventional gas meters in operation. The recent advances in the fields of AI and computer vision offer alternative solutions through intelligent image-based automatic gas meter reading systems. Such systems require huge datasets of gas meter images to train their model on. However, such image datasets are rarely publicly available as the meter images often belong to a service company. Moreover, the few existing public gas meter datasets mostly contain images of the counter meters and not the dials. This limitation would jeopardize the reproducibility and scalability of the proposed automatic gas meter reading systems. Additionally, collecting and annotating a large amount of data is an expensive task restricting research individuals/teams with limited financial resources. To overcome these limitations and in an effort toward supporting the scientific community through open-source open-access initiatives, we introduced a large labeled public benchmark image dataset of residential diaphragm gas meter readings, named the NRC-GAMMA. The dataset contains 28,833 images of the entire gas meter along with 57,766 cropped images of the left and right dial displays, 28,883 images per dial, that were systematically annotated, carefully validated, and intensively monitored/controlled to ensure the quality. The NRC-GAMMA GitHub repository includes metadata, documentation, instructions to use, and scripts to download the archive files. The NRC-GAMMA is an evolving dataset. We are constantly searching for improvements and even more data, therefore, the NRC-GAMMA dataset and metadata will be evolving/growing over time. We recommend that users check the NRC-GAMMA repository for the latest versions of data, metadata, and scripts. We hope the NRC-GAMMA dataset helps the research community in building accurate intelligent solutions to tackle the problem of reading residential meters automatically. This would optimize the currently expensive, labor-intensive, and error-prone manual reading of the gas meters.

References

[1] U.S. Energy Information Administration, "Natural gas and the environment," https://www.eia.gov/energyexplained/natural-gas/natural-gas-and-the-environment.php, 2020, accessed: 12 November 2020.

[2] H. King, "Uses of natural gas," https://geology.com/articles/natural-gas-uses/, 2013, accessed: 12 November 2020.

[3] Government of Canada, "What is in a canadian residential natural gas bill?" https://www.cer-rec.gc.ca/en/data-analysis/energy-commodities/natural-gas/report/canadian-residential-natural-gasbill/index.html, 2020, accessed: 12 November 2020.

[4] Government of Canada, "Provincial and territorial energy profiles," https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/provincial-territorial-energy-profiles/provincial-territorial-energy-profiles-canada.html, 2020, accessed: 12 November 2020.

[5] G. Deconinck, *Metering, Intelligent Enough for Smart Grids?* Dordrecht: Springer, 2010.

[6] C. Son, S. Park, J. Lee, and J. Paik, "Deep learning–based number detection and recognition for gas meter reading." *IEIE Transactions on Smart Processing Computing*, vol. 8, pp. 367–372, 2019.

[7] M. Vanetti, I. Gallo, and A. Nodari, "Gas meter reading from real world images using a multi-net system." *Pattern Recognition Letters*, vol. 34, no. 5, pp. 519–526, 2013.

[8] R. Laroca, Barroso, D. V., M. A., G. R. Gonçalves, W. R. Schwartz, and D. Menotti, "Convolutional neural networks for automatic meter reading." *Journal of Electronic Imaging*, vol. 28, no. 1, p. 013023, 2019.

[9] M. Spichkova, J. van Zyl, S. Sachdev, A. Bhardwaj, and N. Desai, *Comparison of Computer Vision Approaches in Application to the Electricity and Gas Meter Reading.* Springer International Publishing, 2010.

[10] C. P. Edward, "Support vector machine based automatic electric meter reading system." in *Proc. 2013 IEEE International Conference on Computational Intelligence and Computing Research*, 2013.

[11] A. Azeem, W. Riaz, A. Siddique, and U. A. K. Saifullah, "A robust automatic meter reading system based on mask-rcnn." in *Proc. 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, 2020.

[12] F. Yang, L. Jin, S. Lai, X. Gao, and Z. Li, "Fully convolutional sequence recognition network for water meter number reading." *IEEE Access*, 2019.

[13] G. Salomon, R. Laroca, and D. Menotti, "Deep learning for image-based automatic dial meter reading: Dataset and baselines." *arXiv:2005.03106*, 2020.