Care Label Recognition

Jiri Kralicek, Jiri Matas, Michal Busta
Department of Cybernetics
Czech Technical University
Prague, Czech Republic
kralicek@mtit.org, matas@cmp.felk.cvut.cz, bustam@felk.cvut.cz

***Abstract*—The paper introduces the problem of care label recognition and presents a method addressing it. A care label, also called a care tag, is a small piece of cloth or paper attached to a garment providing instructions for its maintenance and information about e.g. the material and size. The information and instructions are written as symbols or plain text.**

Care label recognition is a challenging text and pictogram recognition problem – the often sewn text is small, looking as if printed using a non-standard font; the contrast of the text gradually fades, making OCR progressively more difficult. On the other hand, the information provided is typically redundant and thus it facilitates semi-supervised learning.

The presented care label recognition method is based on the recently published End-to-End Method for Multi-Language Scene Text, E2E-MLT, Busta et al. 2018, exploiting specific constraints, e.g. a care label vocabulary with multi-language equivalences.

Experiments conducted on a newly-created dataset of 63 care label images show that even when exploiting problem-specific constraints, a state-of-the-art scene text detection and recognition method achieve precision and recall slightly above 0.6, confirming the challenging nature of the problem.

***Keywords*-care label recognition, text detection, text recognition, symbol recognition, scene text.**

I. Introduction

Care labels are small tags attached to a garment providing instructions for its care and textual informing about its material, size, etc. The instructions are either expressed as symbols or as terse text, or both, and specify how to appropriately wash, bleach, dry, iron and professionally clean the garment. The textual information is free form, there is no standard. In some economic areas, certain information is mandatory, e.g. the fiber composition in the European Union, but not, for instance, in Switzerland. The care label may also contains other customer and retailer related information such as the size, producer, brand, etc. Part of the text often appears in multiple languages. Care labels are either printed or sewn, or both, see the examples in Fig. 1.

Care label recognition, i.e. the problem of recovering the information and instruction given an image, may facilitate sophisticated robotic garment maintenance units, unsupervised training of material property recognition systems as well as in automated inventory taking.

Care label recognition is an interesting problem. On the one hand, the sewn text looks like a non-standard font. The

Figure 1: Care label examples from the CL2018 dataset. Red rectangles show instruction symbols, orange – textual instructions, yellow – information relevant for care, blue – additional information.

contrast of the text gradually fades, making OCR gradually more difficult. On the other hand, the information and instructions are often redundant, sometimes highly redundant since the text appears in multiple languages besides being expressed in symbols. The redundancy and the simple semantics of the text and symbols open the opportunity for unsupervised learning. We consider the problem in the "wild" setting, where the images of care labels are captured by hand-held devices in natural settings, i.e. not on a scanner.

The care label recognition problem has interesting semantic constraints. The set of materials is limited, the symbols have clear meaning that implies settings of devices like tumble dryers and washing machines. Only certain combination of symbols and materials are permitted. The words related to garment care are from a limited set which can be exploited in care label recognition. In the method described below,

2379-2140/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDAR.2019.00158

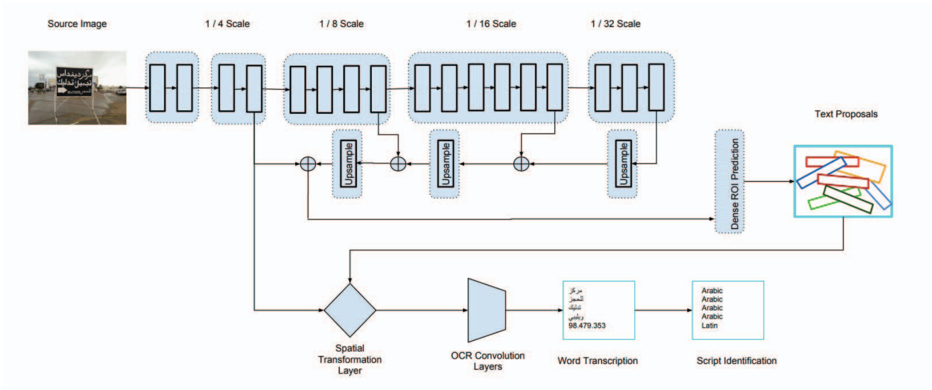


Figure 2: E2E-MLT architecture overview, (courtesy of Busta et al. [1]).

we use a semi-open vocabulary, containing care related words, for care label recognition. The vocabulary contains words from 20 languages; it can be easily extended to other languages and to care-related words such as actions and materials. For further description of the vocabulary, see Sec. IV.

The care label domain is very interesting from the perspective of the multiple language detection and OCR, since it often contains multi-language text in many scripts. Care labels thus provide an easily accessible data source for multi-language detectors and recognition models. A new dataset containing 63 care label images was created, further information is provided in Sec. III.

Another fact worth noting is the variety of different fonts and various text lengths in care labels. The text can be made up of well separated words, but it can be also formed by words without whitespace between them, separated by a delimiter, e.g. "/". From the point of view of the OCR system, this text looks like extremely long words; the detector must thus be able to handle a large range of aspect ratios. Large aspect ratios are challenging even for state-of-the-art models. We adopted the E2E-MLT [1] model, due to its multi-language capability, which we enlarged to care label symbol recognition.

Care label recognition have some similarities with Information extraction from documents [2], [3] and with graphic symbol recognition [4]. Semantic information provided by a symbol and text can be useful in applications such as real world text spotters, where well known symbols such as cross or wheelchair symbol can lead to better recognition of words like "Hospital" or "Reserved parking".

II. The core text detection and recognition method

E2E-MLT [1] with log softmax modification was used as the baseline. E2E-MLT consists of two parts – text localization and text recognition, see Fig. 2. E2E-MLT uses FPN, the feature pyramid network [5] and ResNet-34 [6] as a backbone. The first layer of ResNet-34 is replaced with a sets of 3x3 convolutions with stride 2. The detector works on $\frac{1}{4}$ scale of the original image. The initial convolutional layers are shared between both localization and recognition tasks.

The output prediction consists of 7 channels – per-feature text/non-text confidence, distance from top, bottom, left and right edges of the bounding box containing the pixel and the orientation angle. Dense proposals are filtered by confidence threshold set to 0.9. Locality-aware non-maximum suppression [7] is applied to obtain the final predictions.

The OCR branch is a fully convolutional network. The output log softmax of the OCR module has been modified, which enables multi-language recognition of 7398 characters from six different scripts and of 38 care label symbols. The care label symbols are use runic unicodes, due to missing character equivalents. Images with detected BBox IoU higher than 0.9 were used for training of the OCR part.

The text proposals are used to determine the warp parameters of the spatial transformer layer [8]. The spatial transformer layer normalizes the scale and rotation of the image using bi-linear interpolation to make the learning task of the OCR branch geometry-independent. The input of the OCR is a $\overline{W} \times H \times C$ tensor and the output dimensions are $\frac{\overline{W}}{4} \times \mathcal{A}$. $\overline{W}$ is variable, equal to text width, H is a fixed height set to 40, C is the number of the channels and $\mathcal{A}$ is an alphabet containing concatenation of all characters and care label symbols. $\mathcal{A}$ consists of 7398 characters and 38 symbols.

The overall loss function is a combination of four losses:

$$L_{final} = L_{iou} + \lambda_1 L_{angle} + \lambda_2 L_{dice} + \lambda_3 L_{ctc}, \quad (1)$$

where L_{angle} is a sum over mean squared loss obtained over $\sin(r_\theta)$ and $\cos(r_\theta)$ representations of rotation angle. L_{dice} [9] is used due to high class imbalance between text-regions and background. L_{ctc} proposed in [10] is CTC loss for word-level text recognition. In experiments, we set $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

III. DATA

We created a new dataset of care label images; we are not aware of any such data in the public domain. The dataset consists of 63 care label images captured by hand-held devices in a real environment in different lighting conditions. The datasest is being continuously extended and will be made public. Due to the use of multiple hand-held devices, the dataset consists of images in different resolutions. Each image contains one care label. It is assumed that in applications the care label will be the dominant object, since the image would be acquired with the intention of being suitable for automatic care label processing. Images in the dataset range from simple care labels containing only a minimum of care label symbols and fiber composition in one language to labels with symbols in multiple standards and dense text in multiple languages. The dataset covers many languages and scripts such as Latin, Cyrillic, Greek, CJK.

The annotation includes information about the geometry of the care label data, the type of the data – script such as Latin, Cyrillic etc. or a care label symbol, and the sequence of words or symbols. The care label symbols are annotated by their transcription, e.g.: *washing_not*, see Tab. IV. The geometry is annotated as in the ICDAR 2015 Competition on Robust Reading [11] by coordinates of the four corners of the enclosing convex quadrilateral.

Image are also annotated for care label recognition, with material description – the fiber composition of the garment, and information on bleaching, ironing, tumble drying and washing, such as: "cotton, washing_normal_40, tumble_unknown, ironing_plate_150, bleaching_not". When a care label does not contain information about fiber composition or action, the relevant attribute is declared unknown, e.g. `"ironing_unknown"`.

The dataset is split into evaluation `CL2018_eval` and training sets `CL2018_train`. The `CL2018_train` consists of 32 images, covering many languages, scripts and care label symbol standards. Only Latin, Greek, Cyrillic scripts and the care label symbols according to ISO 3758:2012 [12] were used. The `CL2018_eval` consists of 31 images, covering many languages and care label symbol standards. Only Latin and care label symbols according to ISO 3758:2012 were used. See Fig. 1 for an example. *CL2018* and also new care label dataset *Carelabel128* containing 128 images are available at *http://cmp.felk.cvut.cz/carelabel/*.

IV. METHOD

A. Training

As a baseline, we was used E2E-MLT [1], due to its ability to recognize multi-language text. The model was trained in two steps in an end-to-end fashion. For the initial training, we used a union of the ICDAR RRC-MLT 2017 [13] training dataset, the ICDAR RCTW 2017 [14] training dataset, the ICDAR 2015 [11] training dataset and the Synthetic Multi-language in Natural Scene Dataset [1]. Adam [15] optimizer was used with learning rate = 0.0001, β_1 = 0.9, β_2 = 0.999 and weight decay = 0. In the second step, CL2018_train dataset was used; the learning rate was set to 0.00005. The Model converged in 9900 iterations.

B. The semi-open vocabulary

Care labels contain care-related words such as material, washing and tumble drying process, iron settings, method of the professional cleaning, or bleaching possibilities. The care-related words are a limited set enabling understanding of the care label. In the case of fiber composition, the situation is clear – materials are from a small fixed set whose combination defines the fiber composition. On the other hand, care instructions are not so restricted and the manner highly depends on the manufacturer, e.g. *"Wash cool"* and *"Wash at maximum 30°C"* have the same meaning.

For care label recognition purposes, we created a semi-open vocabulary. The vocabulary includes the set of materials, actions, temperatures and other care related words in many languages often used in the care label domain. The vocabulary will be made public with the paper. The material set consists of five materials – cotton, elastane/lycra, polyester, viscose and wool. The actions describe bleaching, ironing, tumble drying and washing words. The temperatures are associated with an action due to different meaning across the actions. Temperatures can take the form of a word or a number. Other care related words are for example the word *"not"*, which prohibits the action, or the word *"chlorine"*, which is associate with the bleaching process. The semi-open vocabulary contains words from Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Hungarian, Indonesian, Italian, Latvian, Maltese, Norwegian, Polish, Portuguese, Romanian, Spanish, Swedish and Turkish. The semi-open vocabulary consists of 206 words in total and can be easily updated.

C. Text Bounding-Box Prediction Merging

A care label can include dense text regions containing long words. In such cases, a word can be broken into two or more overlapping text bounding-box predictions, see Fig. 3. To improve text bounding-box predictions, we use merging. We assume that the text is mostly oriented horizontally. Text Bounding-Box Prediction Merging has two steps.

In the first step, a list of bounding-box predictions is sorted by the y_2 coordinate and bounding-boxes with y_i coordinates in range of α_1 from the first bounding-box in the list form a line. A bounding-box is removed from the list when assigned to a line. The process is repeated until the list is empty. Bounding boxes in the line are sorted by x_2 coordinate. For further description, see Algorithm 1. In the second step, each bounding-box pair overlapping in x-coordinates by at least α_2 is merged, see Algorithm 2. After initial experiment, α_1 was set to 24 and α_2 was set to 35.

Algorithm 1 Sort bounding-boxes to lines

Input: A list of text bounding-boxes – *BBoxes*, y_{ij} denote j-th y-coordinate of i-th bounding-box, where $j = \{1, 2, 3, 4\}$.
Output: A list of text bounding-boxes sorted in lines

lines $\leftarrow \emptyset$
sort *BBoxes* by y_2
while any *BBox* in *BBoxes* **do**
  for i in range *BBoxes* **do**
    if y_{ij} in range of $y_{1j} \pm \alpha_1$ **then**
      line $\cup \{BBoxes_i\}$
    end if
  end for
  line sort by x_2
  BBoxes $\setminus$ *line*
  append *line* to *lines*
end while
return *lines*

Algorithm 2 Text Bounding-Box Prediction Merging

Input: A list of lines sorted by x_2 coordinates, x_{ij} denote j-th x-coordinate of i-th bounding-box.
Output: A list of merged text bounding boxes

mergedBBoxes $\leftarrow \emptyset$
for *line* in *lines* **do**
  while any *BBox* in *line* **do**
    if only one *BBox* in *line* OR $((x_{13} - \alpha_2) < x_{22}$ OR $(x_{14} - \alpha_2) < x_{21})$ **then**
      merged $= [x_{11}, y_{11}, x_{12}, y_{12}, x_{13}, y_{13}, x_{14}, y_{14}]$
      line $\setminus \{BBox_1\}$
    else
      merged$= [x_{11}, y_{11}, x_{12}, y_{12}, x_{23}, y_{23}, x_{24}, y_{24}]$
      line $\setminus \{BBox_1, BBox_2\}$
    end if
    mergedBoxes $\cup \{merged\}$
  end while
end for
return *mergedBBoxes*

V. Experiments

We conducted three experiments – *Text recognition*, *Symbol recognition* and *Care label recognition*. For evaluation, we used the `CL2018_eval` dataset described in Sec. III. Method input is the E2E-MLT output containing symbols and text longer than 3 characters. An input image resolution is automatically resized to 3696x2048.

A. Text recognition

The goal of the text recognition experiment is to evaluate the care label text in Latin script. Evaluation metrics are: recall R for the text recognition, recall R_{ED1} of the text recognition with edit distance up to 1 from the ground truth, text precision P and text detection D. The quantity R_{ED1} is used since the text often contains a many similar words with the same meaning, e.g. *cotton* and *coton* (a French for cotton), which are important and interchangeable for the care label recognition. During evaluation, the semi-open and dataset vocabulary was used for word correction. Only words with no vocabulary match and with edit distance 1 to any word in vocabulary were corrected.

$$R = \frac{\text{Correctly recognized text}}{\text{All ground-truth text}} \tag{2}$$

$$R_{ED1} = \frac{\text{Recognized text with edit distance up to 1}}{\text{All ground-truth text}} \tag{3}$$

$$P = \frac{\text{Correct text predictions with IoU} > 0.3}{\text{All text predictions}} \tag{4}$$

$$D = \frac{\text{Correct text predictions with IoU} > 0.3}{\text{All ground-truth text}} \tag{5}$$

B. Symbol recognition

The symbol recognition experiment evaluates care label symbols according to ISO 3758:2012 [12]. For symbol evaluation, the following performance measures were used – end-to-end symbol recognition R, precision P and detection D, similarly to (2),(4-5).

C. Care label recognition

The goal of the care label recognition is to recognize the material, bleaching, ironing, tumble drying and washing process. Care label recognition can be text-based, symbol-based or both. When combining both, text-based recognition is preferred due to higher accuracy.

For the care label recognition, we use words from the Semi-open Vocabulary, described in Sec. IV, and the OCR output formed into a sorted list. The list is obtained as follow: the OCR output is sorted by Algorithm 1; the output of the sorting algorithm is a set of sorted lines by y_2-coordinate containing text and symbols sorted by x_2-coordinate; the list is formed by the text and symbols according to their order and in the experiment is referred as a text.

Symbol-based recognition is the same for each subtask – bleaching, ironing, tumble drying, washing. When any care label symbol related to the subtask occurs, its counter is increased and is used for a recognition. If more than one symbol is recognized, the symbol with the highest value of the counter will be selected. Text-based recognition is subtask dependent.

1) Material: Material recognition is focusing on the fiber composition of the garment, describing materials forming the garment and is represented by individual materials without their percentage, separated by backslash, e.g.

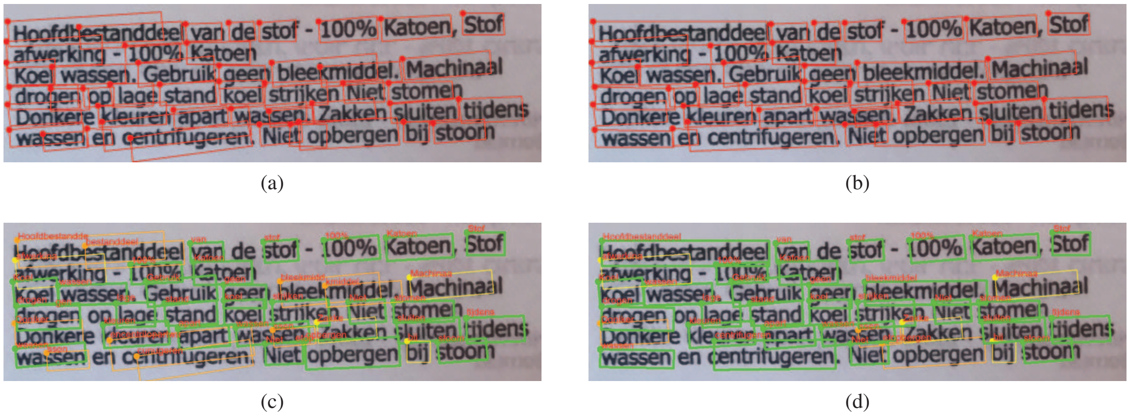


Figure 3: Text box predictions and OCR output before (a,c) and after (b,d) Text Bounding-Box Prediction Merging. Color coding of the OCR output: green – correct recognition, yellow – recognition with edit distance 1, orange – recognition with edit distance > 1.

"wool/elastane". Recognition is purely text-based, because there are no symbols specifying fiber composition.

The text is searched for material using material words obtained from the Semi-open vocabulary. Any material that is found at least once is added to the recognition. We recognize five materials – *cotton, elastane/lycra, polyester, viscose* and *wool*. If none of the materials is found in the text, material recognition will output *unknown_material*. See examples in Tab. V.

Material recognition suffers in images where information is provided by only one word. In these cases, many material words are recognized with edit distance 1 or split to two text predictions. Use of vocabulary and Text Bounding-Box Prediction Merging may improve performance, see Tab. III

Table I: Text Recognition

End-to-end text recognition results on CL2018 evaluation dataset. Recall (R), Recall with edit distance up to 1 (R_{ED1}), Precision (P), Detection(D), Text Bounding-Box Prediction Merging (Merged Predictions), Semi-open Vocabulary (SOV), Dataset Vocabulary(DV).

Method	**R**	$\mathbf{R_{ED1}}$	**P**	**D**
Baseline	0.391	0.636	0.590	0.942
+ Merge Prediction	0.408	0.643	0.718	0.933
+ SOV	0.419	0.632	0.590	0.942
+ Merged Predictions, SOV	0.436	0.642	0.718	0.933
+ DV	0.589	0.635	0.590	0.942
+ Merged Predictions, DV	0.599	0.642	0.718	0.933

Table II: Symbol Recognition

End-to-end symbol recognition results on CL2018 evaluation dataset. Recall (R), Precision (P), Detection(D), Text Bounding-Box Prediction Merging (Merged predictions)

Method	**R**	**P**	**D**
Baseline	0.422	0.436	0.750
+ Merged predictions	0.375	0.489	0.688

2) Bleaching: Bleaching recognition is a combination of text-based and symbol-based recognition and it is represented by the bleaching process. There are three possible bleaching processes – *Do not bleach*, *Non-chlorine bleach* and *Any bleaching agent*, see Tab. IV.

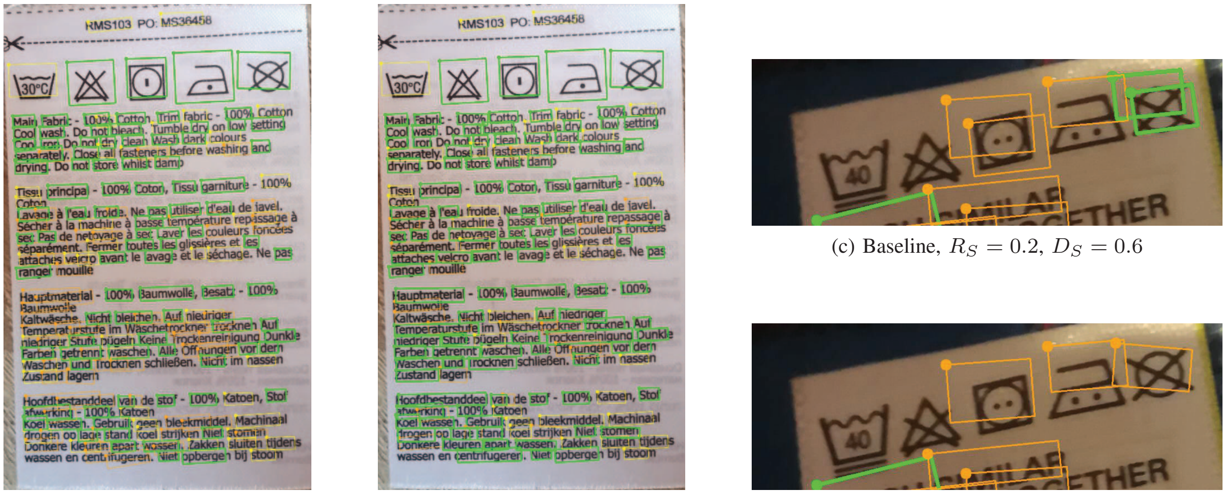
Text-based bleaching recognition is based on *bleaching*, *chlorine* and *not* words, which have to occur in the distance D from each other. D is set to 6. *Do not bleach* counter is increased when the only words detected are *bleaching* and *not*. *Non-chlorine bleach* requires the words not and *chlorine*. When only the word *bleach* occurs in the text, *Any bleaching agent* counter is increased. The process with the highest counter value is selected. When bleaching is not recognized, the output is *bleaching_unknown*.

If the text contain *bleaching* and *temperature/not* words in distance d, counter of the temperature or not word is increased. d is set to 6. The temperature or not modification is selected by the highest counter value.

3) Ironing: Ironing recognition is a combination of text-based and symbol-based recognition and it is defined by the temperature. There are three possible temperatures – *Low*, *Medium* and *Hot*, and *Not* modification prohibiting ironing, see Tab. IV.

Text-based ironing recognition is based on the *ironing*, *temperature* and *not* words obtained from the Semi-open Vocabulary. When the text contains *ironing* and *temperature/not* words in distance d, counter of the *temperature* or *not word* is increased. d is set to 6. The *temperature* or *not* modification is selected by the highest counter value. When ironing is not recognized, the output is *ironing_unknown*.

4) Tumble drying: Tumble drying recognition is a combination of text-based and symbol-based recognition and it is represented by the tumble drying cycle. There are three tumble drying cycles – *Normal process*, *Mild process* and *Do not tumble dry*, see Tab. IV.



(a) Baseline, $R_T = 0.591$
(b) Merged, $R_T = 0.691$
(c) Baseline, $R_S = 0.2$, $D_S = 0.6$
(d) Merged, $R_S = 0$, $D_S = 0.6$

Figure 4: Examples of text and symbol recognition. Text recall (R_T), Symbol recall (R_S), Symbol detection (D_S), Text Bounding-Box Prediction Merging (Merged). Text recognition color coding: green – correct recognition, yellow – recognition with edit distance 1, orange – recognition with edit distance > 1. Symbol recognition color coding: green – correct recognition, orange – correct detection.

Table III: Care label recognition accuracy on the CL2018 evaluation dataset

Fiber composition (Material), Text Bounding-Box Prediction Merging (Merged Predictions).

Method	Material	Bleaching	Ironing	Tumble drying	Washing
Baseline	0.839	0.645	0.613	0.742	0.548
+ Merged Predictions	0.806	0.677	0.581	0.774	0.516
+ Semi-open Vocabulary	0.484	0.677	0.645	0.710	0.581
+ Merge Predictions, Semi-open Vocabulary	0.677	0.677	0.645	0.710	0.548
+ Dataset Vocabulary	0.935	0.677	0.645	0.710	0.613
+ Merge Predictions, Dataset Vocabulary	0.968	0.677	0.645	0.742	0.581

Text-based recognition is based on *tumble drying*, *temperature*, *process* and *not* words. Temperature and process are equivalent, therefore both are referred as a temperature. When the text contain *tumble drying* and *temperature/not* word in distance d, the counter of the *temperature* or *not* word is increased. d is set to 6. Cycle with highest value is chosen. When no tumble drying cycle is recognized, the output is *tumble_unknown*.

5) Washing: Washing recognition is a combination of text-based and symbol-based and is represented by wash cycle. Wash cycle specify washing process and maximum temperature, e.g.: *"washing_veryMild_40"*. There are four possible temperatures – *30°C, 40°C, 60°C, 95°C*, and four processes – *normal, mild, very mild* and *by hand*. Eleven wash cycles are recognized in total, see Tab. IV.

For washing recognition, the text is searched for the wash cycle symbol and for the *wash*, *temperature* and *not* words, obtained from the Semi-open Vocabulary.

Text-based washing recognition distinguishes only *temperature* and *not* modification, which prohibits washing, and it assumes *normal* process, because there is no occurrence of process type in the evaluation dataset text. However, the code can be easily extended by process types. When text contain *wash* and *temperature/not* word at a distance d, the counter of the *temperature*, or *not* word is increased. d is set to 6. Temperature with highest value is chosen and *not* modification is preferred. When there is no wash cycle recognized, the output is *"washing_unknown"*.

Material recognition is often corrupted due to the word *wool*, which is a part of the word *cotton* in many languages such as *"baum-wolle"* (German), *"ba-wełna"* (Polish) and "ba-vlna" (Czech), where the second part of the word means *wool*. Many care labels contain instructions only in a symbol form without any redundancy and care-related words such as temperature and process, are in some cases split to multiple text bounding-boxes predictions or recognized with edit distance 1, which leads to worse results compare to the material recognition.

Although Text Bounding-Box Prediction Merging leads to worse symbol results, combined with vocabulary provides the best results, due to textual redundancy.

Table IV: Symbol – Transcription

No.	Symbol	Transcription	No.	Symbol	Transcription	No.	Symbol	Transcription
0		unspecified material/bleaching/ironing/tumble drying/washing						
1		bleaching_any	14		drying_tumble_not	27		professional_wet_veryMild_W
2		bleaching_not	15		ironing_not	28		washing_hand_40
3		bleaching_onlySome	16		ironing_plate_110	29		washing_mild_30
4		drying_flat	17		ironing_plate_150	30		washing_mild_40
5		drying_flat_drip	18		ironing_plate_200	31		washing_mild_60
6		drying_flat_drip_shade	19		professional_dry_mild_F	32		washing_normal_30
7		drying_flat_shade	20		professional_dry_mild_P	33		washing_normal_40
8		drying_line	21		professional_dry_normal_F	34		washing_normal_60
9		drying_line_drip	22		professional_dry_normal_P	35		washing_normal_95
10		drying_line_drip_shade	23		professional_dry_not	36		washing_not
11		drying_line_shade	24		professional_wet_mild_W	37		washing_veryMild_30
12		drying_tumble_lower_60	25		professional_wet_normal_W	38		washing_veryMild_40
13		drying_tumble_normal_80	26		professional_wet_not			

VI. Conclusion

The paper introduced the problem of a care label recognition. A care label recognition method based on the E2E-MLT text detection and recognition algorithm was presented and evaluated on a newly-created care label dataset. The dataset will be made public. Experiments show that the care label domain is challenging even for state-of-the-art models. The proposed method achieved the following care label recognition rates: fiber composition – 96.8%, bleaching – 67.7%, ironing – 64.5%, tumble drying – 74.2% and washing – 58.1%.

References

[1] Y. Patel, M. Busta, and J. Matas, "E2E-MLT - an unconstrained end-to-end method for multi-language scene text," *CoRR*, vol. abs/1801.09919, 2018.

[2] X. Zhao, Z. Wu, and X. Wang, "Cutie: Learning to understand documents with convolutional universal text information extractor," 2019.

[3] M. Holecek, A. Hoskovec, P. Baudi, and P. Klinger, "Line-items and table understanding in structured documents," 2019.

[4] M. Luqman, T. Brouard, and J.-Y. Ramel, "Graphic symbol recognition using graph based signature and bayesian network classifier," *ICDAR*, 2010.

[5] T.-Y. Lin, P. Dollár, R. B. Girshick *et al.*, "Feature pyramid networks for object detection," *CVPR*, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CVPR*, 2016.

[7] X. Zhou, C. Yao, H. Wen *et al.*, "East: An efficient and accurate scene text detector," *CVPR*, 2017.

[8] M. Jaderberg, K. Simonyan, A. Zisserman *et al.*, "Spatial transformer networks," in *NIPS*, 2015.

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *3DV*, 2016.

[10] A. Graves, S. Fernández, F. J. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *ICML*, 2006.

[11] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou *et al.*, "ICDAR 2015 competition on robust reading," in *ICDAR*, 2015.

[12] Ginetex, "Textile care symbols." 2018. [Online]. Available: http://www.ginetex.net/userfiles/files/Textile_care_symbols_en.pdf

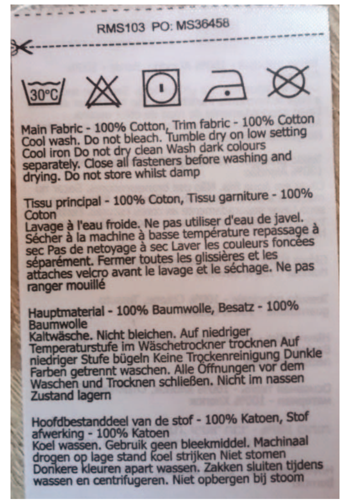
[13] N. Nayef, F. Yin, I. Bizid *et al.*, "ICDAR2017 robust reading challenge on multi-lingual scene text detection and script identification - RRC-MLT," in *ICDAR*, 2017.

[14] B. Shi, C. Yao, M. Liao *et al.*, "ICDAR2017 competition on reading chinese text in the wild (RCTW-17)," in *ICDAR*, 2017.

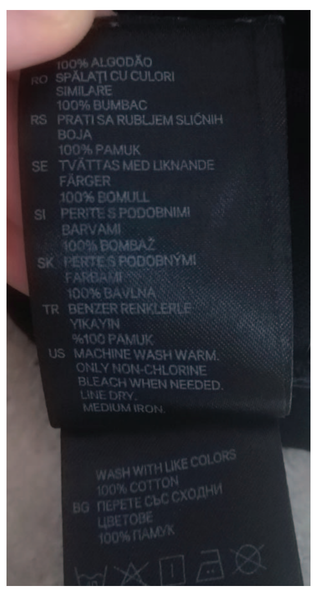
[15] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ICLR*, 2014.

Table V: Care label recognition examples

The images were cropped for presentation purposes. Incorrect recognition is highlighted in red.



GT	Recognized	GT	Recognized	GT	Recognized
unknown_material	unknown_material	cotton/viscose	cotton/viscose	cotton	cotton
washing_normal_30	washing_normal_30	washing_mild_40	washing_normal_30	washing_normal_30	washing_normal_30
tumble_unknown	tumble_unknown	drying_tumble_not	drying_tumble_not	drying_tumble_lower_60	drying_tumble_not
ironing_plate_150	ironing_plate_150	ironing_plate_110	ironing_plate_110	ironing_plate_110	ironing_plate_110
bleaching_onlySome	bleaching_onlySome	bleaching_not	bleaching_not	bleaching_not	bleaching_not



GT	Recognized	GT	Recognized	GT	Recognized
cotton	cotton	cotton	cotton	cotton	cotton
washing_mild_40	washing_normal_30	washing_normal_40	washing_normal_40	washing_normal_40	washing_not
drying_tumble_not	drying_tumble_not	tumble_unknown	drying_tumble_normal_80	drying_tumble_normal_80	drying_tumble_normal_80
ironing_plate_110	ironing_plate_110	ironing_plate_150	ironing_plate_150	ironing_unknown	ironing_plate_150
bleaching_not	bleaching_not	bleaching_unknown	bleaching_not	bleaching_onlySome	bleaching_unknown