Journal of Universal Computer Science, vol. 29, no. 5 (2023), 419-431
submitted: 11/6/2022, accepted: 19/1/2023, appeared: 28/5/2023

Semi-Supervised Semantic Segmentation for Identification of Irrelevant Objects in a Waste Recycling Plant

César Domínguez
(Universidad de La Rioja, Logroño, La Rioja, Spain
https://orcid.org/0000-0002-2081-7523, cesar.dominguez@unirioja.es)

Jónathan Heras
(Universidad de La Rioja, Logroño, La Rioja, Spain
https://orcid.org/0000-0003-4775-1306, jonathan.heras@unirioja.es)

Eloy Mata
(Universidad de La Rioja, Logroño, La Rioja, Spain
https://orcid.org/0000-0003-0538-4579, eloy.mata@unirioja.es)

Vico Pascual
(Universidad de La Rioja, Logroño, La Rioja, Spain
https://orcid.org/0000-0003-3576-0889, mvico@unirioja.es)

Lucas Fernández-Cedrón
(SpectralGeo, Logroño, La Rioja, Spain
lucas@spectralgeo.es)

Marcos Martínez-Lanchares
(SpectralGeo, Logroño, La Rioja, Spain
marcos@spectralgeo.es)

Jon Pellejero-Espinosa
(SpectralGeo, Logroño, La Rioja, Spain
jon@spectralgeo.es)

Antonio Rubio-Loscertales
(SpectralGeo, Logroño, La Rioja, Spain
antonio@spectralgeo.es)

Carlos Tarragona-Pérez
(SpectralGeo, Logroño, La Rioja, Spain
carlos@spectralgeo.es)

Abstract: In waste recycling plants, measuring the waste volume and weight at the beginning of the treatment process is key for a better management of resources. This task can be conducted by using orthophoto images, but it is necessary to remove from those images the objects that are not involved in the measurement process such as containers or trucks. This work proposes the application of deep learning for the semantic segmentation of those irrelevant objects. Several deep architectures are trained and compared, while three semi-supervised learning methods (PseudoLabeling, Distillation and Ensemble Distillation) are proposed to take advantage of non-annotated images. In these experiments, the U-net++ architecture with an EfficientNetB3 backbone, trained with the set of labelled images, achieves the best overall multi Dice score of 91.48%. The application of semi-supervised learning methods further boosts the segmentation accuracy in a range between 1.82% and 3.92%, on average.

Keywords: Waste management, Deep Learning, Semi-Supervised Learning, Semantic Segmentation, Orthophoto
Categories: I.2, I.4
DOI: 10.3897/jucs.87643

1 Introduction

As nations and cities become more populated and prosperous, offer more products and services to citizens, and participate in global trade and exchange, they face corresponding amounts of waste to manage through treatment and disposal. By 2050, the world is expected to generate 3.40 billion tons of waste annually, increasing drastically from today's 2.01 billion tons [Kaza et al., 2018]. Therefore, efficient recycling strategies are critical to reduce the devastating environmental effects of rising waste production [Bashkirova et al., 2021]. In this context, waste recycling plants are central since, in these plants, the collected recyclable waste is sorted into separate bales of plastic, paper, metal and glass.

In order to achieve a better management of resources in waste recycling plants, a key indicator is the waste volume and weight at the beginning of the treatment process. This measurement can be performed by using orthophoto images (that is, aerial images that have been geometrically corrected such that their scale is uniform and true distances can be measured) [Ortenzi et al., 2021]; however, it is necessary to process those images to identify and discard objects (like containers or trucks) that might appear in the image, but should not be taken into account in the measurement process — this task is known as object removal, and plays a key role as a pre-processing step to measure properties of objects in images [Pally and Samadi, 2022]. This issue can be faced by means of semantic segmentation algorithms that serve to classify every pixel of an image among target classes of interest [Gonzalez et al., 2002]. Currently, semantic segmentation tasks are mainly tackled by using deep learning methods [LeCun et al., 2015].

Deep learning has many applications in waste management including waste classification [Huang et al., 2020, Meng and Chu, 2020], waste object localisation in outdoor scenarios [Sousa et al., 2019, Proença and Simões, 2020], waste detection and segmentation in materials recovery facilities [Bashkirova et al., 2021], or recognising composition of construction waste mixtures [Lu et al., 2022]. Deep learning methods have been recently used for classifying and detecting objects in waste recycling facilities. For instance, in [Yang et al., 2022], the YOLO object detection algorithm was used to detect electrical and electronic equipment that contain lithium batteries since they have to be processed differently on waste disposal plants; and, two classifiers were combined to classify recyclables and distinguish types of plastics in [Vogiatzis et al., 2021]. There are also a few works that deal with semantic segmentation tasks in waste recycling facilities. Namely, [Bchir et al., 2021] employed a DeepLabv3+ model to segment Polyethylene Terephthalate objects automatically; and, [Sievert, 2021] conducted a comparison of instance segmentation models to guide unmanned vehicles for autonomous litter collection. Finally, the Visual Domain Adaptation 2022 Challenge was recently released with the aim of developing models for automatically industrial waste sorting [Bashkirova et al., 2022].

However, and up to the best of our knowledge, deep learning methods have not been used for segmenting objects in actual waste recycling plants. One of the main challenges for successfully applying deep learning methods is the necessity of a great amount of images that must be manually annotated. Such an annotation process is a tedious and time-consuming task that can take several hours or even days [Lin et al., 2014, Li et al., 2020]. In order to reduce such a burden, close-transfer learning [Razavian et al., 2014] and semi-supervised methods [Zhu and Goldberg, 2009] can be applied. The former methods use the knowledge learned on a close task where acquiring images in easier than in the final task; whereas, the latter methods take advantage of both labelled and unlabelled data. These two approaches have been studied in this work. Namely, we are focused on combining close-transfer techniques and semi-supervised learning methods with deep learning models to produce the exact segmentation of objects that appear in orthophotos of recycling plants, but that should be removed to precisely measure waste volumes. The original contribution of this paper is threefold:

– The analysis of a close-domain transfer learning approach and three semi-supervised learning models to deal with the small size of the annotated dataset by taking advantage of raw images and unlabelled orthophoto images.

– A detailed comparison of several state-of-the-art deep neural networks for semantic segmentation (both architectures and backbones) for processing orthophoto images.

– A statistical analysis to identify whether there are significant differences among the studied deep learning models and the semi-supervised learning methods.

As a result, this paper demonstrates that using a semi-supervised learning technique allows us to successfully train segmentation models, substantially reducing the effort required to annotate many images. Then, the comparison of the networks shows that the U-net++ architecture with the EfficientNetB3 backbone achieves the best performance in segmentation accuracy. In this case, the multi Dice score is equal to 91.65%. This result confirms that orthophoto images can be effectively processed to segment the environment of recycling plants and find objects of interest. This step will enable the use of orthophoto images for measuring waste volume with a higher precision.

The rest of the paper is organised as follows: the first section on Materials and Methods describes the input datasets, the semantic segmentation models, the semi-supervised learning methods and the way results are evaluated and compared; the Experimental Result section presents the outcomes of the tests; and, the last section ends the manuscript with final comments and remarks on future activities.

2 Materials and methods

2.1 Input dataset

This paper tackles the problem of image segmentation from orthophoto images captured in a recycling plant. Within these lines, the automatic segmentation of orthophoto images is achieved by representing them in more descriptive and discriminative feature spaces, learned from actual images, where pixels having similar semantic attributes can be grouped and labelled in different classes. A set of annotated images is thus required to allow the training of the models. At the same time, additional annotated images are

Figure 1: Sample orthophoto image and corresponding annotated image. black pixels belong to the class person, yellow to the container class, red to the forklift class, green to the truck class, and black to background

needed to evaluate the classification results on a ground truth. These two sets of images (training and test sets) form the annotated dataset.

The annotated dataset is made of 49 manually-labelled colour images acquired by a Safire IP 8MPx camera with a focal of 2.8mm in a recycling plant in Spain. In order to construct the orthophotos, 10 images of resolution 2560×1440 pixels were taken and combined using the ODM technology through the PyODM library [OpenDroneMap project, 2021]. The orthophoto images have a resolution of 1526×1468 pixels, and were manually annotated using the Labelme tool [Wada et al., 2021] as shown in Figure 1. The annotation aims to separate five classes of interest: person (black segments), container (yellow segments), forklift (red segments), truck (green segments), and background (black segments). All the images were taken from the same facility, but there is a considerable variability in the images since they are taken on different days and with different amounts of waste. Namely, the number and position of containers in the images varies from image to image; besides, the amount of waste in them changes from image to image. Similarly, the number of people and their position changes from image to image. Finally, the two classes of objects with most variability are forklift and truck since the generation of the orthophotos deforms those objects in the images.

The manual annotation of images is a time-demanding and tedious task. Although the acquisitions provide many images, the annotation has been limited to the representative subset of 49 images described previously. However, there are 322 further orthophoto images, not-labelled but acquired under the same experimental conditions. These orthophoto images will tune the training of the networks through the implementation of three semi-supervised approaches. The network architectures and the semi-supervised algorithms will be detailed in the following subsection.

2.2 Semantic segmentation models

As stated in the previous section, the 49 labelled orthophoto images, randomly splitted into training sets (39 images) and test sets (10 images) using a 5-fold cross-validation approach, were used to set up and evaluate the deep segmentation architectures (see Table 1 for the number of objects in each dataset). From the training set, several deep-learning segmentation algorithms were fine-tuned [Razavian et al., 2014]. Namely, 7 architectures

	Person	Container	Forklift	Truck
Training set 1	34	296	38	10
Test set 1	8	75	9	2
Training set 2	32	297	38	9
Test set 2	10	74	9	3
Training set 3	32	293	38	11
Test set 3	10	78	9	1
Training set 4	32	292	37	11
Test set 4	10	79	10	1
Training set 5	38	306	37	7
Test set 5	4	65	10	5

Table 1: Number of objects of interest in the training and test sets

were trained, they are summarised in Table 2 — we fixed a seed for reproducibility and train each model just once. For training, we used the libraries PyTorch [Paszke et al., 2019] and FastAI [Howard and Gugger, 2020]; and using a GPU Nvidia RTX 2080 Ti. The procedure presented in [Howard and Gugger, 2020] was employed to set the learning rate for the different architectures, the learning rate for the first layers of the models was fixed to 1e-4, and for the last layers of the models to 1e-3. Also, early stopping was applied in all the architectures to avoid overfitting (validation loss was monitored and the training process stopped when such a validation loss did not improve after 5 epochs). As a result of the training process, several models were produced that can be used for inference by providing them a natural image as input. Then, the models will output the mask associated with the segmentation.

In addition, we have applied a close transfer-learning approach for training our models. When applying transfer learning, it is well-known the importance of using a source task that is as close as possible to the target task [Mensink et al., 2021]. Therefore, we have used 404 raw images that were used to construct the orthophotos of the training set (390 images were used to generate the 39 ortophoto images of the training dataset, and the other 14 raw images were introduced to increase the variability of the dataset, but were not used for generating orthophoto images). These 404 images were manually annotated, and used to train the models from Table 2 — annotating raw images in easier than annotating orthophoto images due to the distortions that might appear in the latter images. Subsequently, those models were used as starting point to train the same models but using the orthophoto images.

Using a 5-fold cross-validation approach, all the models were then evaluated on the test set of 10 annotated orthophoto images using the multi-class Dice score [Opitz and Burst, 2019]. This metric is defined using the precision, P_i, and recall, R_i, values for each class defined as:

$$P_i = \frac{m_{ii}}{\sum_{x=1}^{n} m_{ix}}; \quad R_i = \frac{m_{ii}}{\sum_{x=1}^{n} m_{xi}}$$

where n is the number of classes, and m_{jk} for $j = 1, \ldots, n$ and $k = 1, \ldots, n$ is the total number of pixels predicted as $class_j$, whose actual label is $class_k$. From the precision

Architecture	Backbones
Bisenet	Resnet18, Resnet34
Deeplabv3+	Resnet50, Resnext50, EfficientNetB3
HRNet	w30
Manet	Resnet50, Resnext50, EfficientNetB3
PAN	Resnet50, Resnext50, EfficientNetB3
U-net	Resnet50, Resnest50, EfficientNetB3
U-net++	Resnet50, Resnest50, EfficientNetB3

Table 2: Segmentation architectures and the backbones employed in this work

and recall values, the multi-class Dice score is defined as follows:

$$MultiDice = \frac{1}{n} \sum_{i=1} \frac{2P_i R_i}{P_i + R_i}$$

2.3 Semi-supervised learning methods

In order to take advantage of the unlabelled images, 3 semi-supervised learning approaches were employed. Namely, we have employed PseudoLabeling [Lee, 2013], Distillation [Hinton et al., 2015], and Ensemble Distillation [Bucila et al., 2006] — the latter method is also known as Model Distillation.

The PseudoLabeling approach consists of two steps; first, we employ a model trained on a manually labelled dataset to make predictions in an unlabelled dataset; secondly, the manually and automatically-labelled datasets are combined to train a new model using the same architecture employed in the original model. We have applied the PseudoLabeling approach to all the architectures presented in the previous section; and, the initial model was trained with the close-transfer learning approach.

The Distillation approach is similar to the PseudoLabeling approach, but in the second step, the model might have a different underlying architecture than the model employed for the first step. In our case, we have trained several models using the training procedure presented in the previous section, and selected the best model for generating the automatically-labelled dataset. Furthermore, we have used the combination of the manually and automatically-labelled datasets to train all the architectures presented in the previous section.

Finally, Ensemble Distillation differs from the Distillation approach in the way of producing the automatically labelled dataset; namely, instead of using a single model for making predictions in an unlabelled dataset, the predictions are generated from an ensemble of models. In this work, we have employed the 5 models with the highest total multi Dice score for producing the predictions on the unlabelled dataset; and, as in the previous approaches, the manually and automatically-labelled datasets were used to train all the architectures presented in the previous section.

3 Experimental Results

The performance of the trained networks (both by applying and without applying the semi-supervised learning methods) was evaluated using a 5-fold cross-validation approach

'RPtQJXH]&   +HUDV-   0DWD(3DVFXDO9   )HUQiQGH]/   0DUWtQH]0

0RGHO	%DVHOLQH	&ORVH	7UDQVIHU	3VHXGRODEHO	'LVWLOODW
%LVHQHW 5HVQHW					
%LVHQHW 5HVQHW					
'HHSODEY 5HVQHW					
'HHSODEY 5HVQH[W					
'HHSODEY (IILFLHQW1HW%					
+51HW Z					
0DQHW 5HVQHW					
0DQHW 5HVQHVW					
0DQHW (IILFLHQW1HW%					
3$1 5HVQHW					
3$1 5HVQH[W					
3$1 (IILFLHQW1HW%					
8 QHW 5HVQHW					
8 QHW 5HVQHVW					
8 QHW (IILFLHQW1HW%					
8 QHW 5HVQHW					
8 QHW 5HVQHVW					
8 QHW (IILFLHQW1HW%					

7DEOH   0HDQ VWG PXOWL 'LFH VFRUH IURP WKH DSSOLFDWLR
SURFHGXUHV 7KH UHVXOW LQ EROG LV WKH EHVW

ZLWK LQGHSHQGHQW WHVW VHWV WKDW FRQVLVW RI   RUWKRSK
$OO WKH PRGHOV WUDLQHG RQO\ ZLWK WKH PDQXDOO\ ODEH
QRZ RQ ZH ZLOO UHIHU WR WKLV DSSURDFK DV EDVHOLQH DFK
DQG RQO\ WKH PRGHO WUDLQHG ZLWK WKH 8 QHW   DUFKLWH
EDFNERQH REWDLQHG D PXOWL 'LFH VFRUH RYHU   VHH 7DEOH
7KH UHVXOWV FDQ EH FRQVLGHUDEO\ LPSURYHG E\ DSSO\LQJ
DSSURDFK VHH 7DEOH   1DPHO\ VRPH PRGHOV LPSURYHG PRU
HUDJH D   ,I WKH VHJPHQWDWLRQ QHWZRUNV DUH FRPSDUH
'HHSODEY   5HVQHW   'HHSODEY   (IILFLHQW1HW% 8 QHW (I
5HVQHVW DQG 8 QHW (IILFLHQW1HW% ZLWK D WRWDO PXOWL
WKHP WKH 8 QHW DUFKLWHFWXUH WRJHWKHU ZLWK WKH (IILFLH
VHJPHQWDWLRQ DFFXUDF\ WKDQ WKH RWKHU QHWZRUNV 1DPH
'LFH VFRUH RI   DQG WKH LPSURYHPHQW UHJDUGLQJ LWV EDV
LQ)LJXUH
7KH LPSDFW RI WKH GLIIHUHQW VHPL VXSHUYLVHG OHDUQLQJ PH
LV DOVR VKRZQ LQ 7DEOH :LWK PRUH GHWDLOV WKH 'LVWLOOD
LPSURYHPHQW RI   ZLWK D VWDQGDUG GHYLDWLRQ RI   2Q
UHVXOWV XVLQJ WKLV WUDLQLQJ DSSURDFK ZKLOH LQ VRPH FDVH
PRGHO WKH LPSURYHPHQW LV RYHU   VHH)LJXUH   6LPLODUO\
SURGXFHV D PHDQ LPSURYHPHQW RI   ZLWK D VWDQGDUG G
VHYHQ QHWZRUNV KDYLQJ ZRUVH UHVXOWV)LQDOO\ WKH (QV
FRQVLGHUDEO\ LPSURYHV WKH SHUIRUPDQFH RI WKH PRGHOV D P
GHYLDWLRQ RI
,Q RXU DQDO\VLV ZH FDQ VHH WKDW WKHUH DUH VRPH DUFKLW
EHWWHU WKDQ WKH RWKHUV LQGHSHQGHQWO\ RI WKH WUDLQLQJ P
ZRUVH UHVXOWV ZHUH WZR DUFKLWHFWXUHV EDVHG RQ WKH DWW
@ DQG 0DQHW >/L HW DO   @ 7KLV NLQG RI QHWZRUN UHT

D E

F G

)LJXUH ([DPSOH RI WKH VHJPHQWDWLRQ UHVXOWV XVLQJ WKH 8
WUDLQHG ZLWK WKH EDVHOLQH DQG FORVH WUDQVIHU DSSURDF
WUXWK F %DVHOLQH PRGHO G &ORVH WUDQVIHU P

WUDLQHG SURSHUO\ >'RVRYLWVNL\ HW DO @ DQG XQIRUW
VPDOO GDWDVHW 7KH RWKHU DUFKLWHFWXUHV DUH EDVHG RQ F
REWDLQHG EHWWHU UHVXOWV KRZHYHU ZH FDQ DOVR QRWLFH V
%LVHQHW PRGHOV REWDLQHG ZRUVH UHVXOWV GXH WR WKH VL]H
5HVQHW WKDW DUH VPDOOHU WKDQ RWKHU EDFNERQHV VXFK D
WKH FDVH RI WKH 8 QHW PRGHOV WKH 8 QHW YHUVLRQ DFKLHYH
WKH VHYHUDO LPSURYHPHQWV LQWURGXFHG LQ WKLV QHZ YHUVL
HW DO @)LQDOO\ WKH 'HHS/DEY +51HW DQG 8 QHW R
LQGHSHQGHQWO\ RI WKH VHOHFWHG EDFNERQH

,Q DGGLWLRQ WR VHDUFKLQJ IRU WKH EHVW SHUIRUPLQJ PRGHO
VWXG\ WR GHWHUPLQH ZKHWKHU WKH UHVXOWV REWDLQHG ZLWK W
VWDWLVWLFDOO\ VLJQLILFDQW 7R WKLV DLP VHYHUDO QXOO K\S
LQJ WKH PHWKRGRORJ\ SUHVHQWHG LQ >*DUFLD HW DO 6KHV
EHWZHHQ D SDUDPHWULF RU D QRQ SDUDPHWULF WHVW WR FRP
FRQGLWLRQV LQGHSHQGHQFH QRUPDOLW\ DQG KHWHURVFHG
WHVW LV RQO\ DSSURSULDWH ZKHQ WKH WKUHH FRQGLWLRQV DU

,Q WKLV VWXG\ WKH LQGHSHQGHQFH FRQGLWLRQ LV IXOILOOH
OHDUQLQJ DSSURDFK LV LQGHSHQGHQW RI WKH RWKHUV :H XVH W
DQG :LON @ WR FKHFN QRUPDOLW\ ² ZLWK WKH QXOO K\SRWKHV
D QRUPDO GLVWULEXWLRQ ² DQG D /HYHQH WHVW >/HYHQH @
² ZLWK WKH QXOO K\SRWKHVLV EHLQJ WKDW WKH UHVXOWV DUH K

D E

F G

H I

)LJXUH ([DPSOH RI WKH VHJPHQWDWLRQ UHVXOWV XVLQJ WKH
DSSURDFKHV ZLWK WKH 8 QHW 5HVQHW PRGHO D ,QSXW LPD
&ORVH WUDQVIHU PRGHO G 'LVWLOODWLRQ H 3VHXGR/DEHO

PRUH WKDQ WZR WUDLQLQJ DSSURDFKHV ZH ZLOO HPSOR\ DQ $
FRQGLWLRQV DUH IXOILOOHG DQG D)ULHGPDQ WHVW RWKHUZLVH
QXOO K\SRWKHVLV ZLOO EH WKDW DOO WKH PRGHOV KDYH WKH VD
FKHFNLQJ ZKHWKHU D PRGHO LV VWDWLVWLFDOO\ EHWWHU WKDQ
SURFHGXUH LV HPSOR\HG WR DGGUHVV WKH PXOWLSOH K\SRWK
PRGHOV $ +ROP SRVW KRF SURFHGXUH >+ROP @ LQ WKH Q
%RQIHUURQL 'XQQ SRVW KRF SURFHGXUH >6KHVNLQ @ LQ WK
GHWHFWLQJ VLJQLILFDQFH RI WKH PXOWLSOH FRPSDULVRQV >*D
DQG p YDOXHV VKRXOG EH FRUUHFWHG DQG DGMXVWHG :H KDYH S

DQDO\VLV ZLWK D OHYHO RI 0.05 FRQILGHQFH WKH QXOO K\SRWKHVLV WKDW WKH HIIHFW KD
PHDVXUHG XVLQJ &RKHQ¶V G >&RKHQ @ DQG (WD 6TXDUHG >&
,Q RXU VWXG\ DQG VLQFH WKH QRUPDOLW\ FRQGLWLRQ ZDV QRW
: $p < 0.001$ D)ULHGPDQ¶V QRQ SDUDPHWULF WHVW ZDV HPS
WKH WUDLQLQJ SURFHGXUHV)ULHGPDQ¶V WHVW SHUIRUPHG D U
XQGHU FRPSDULVRQ VHH 7DEOH DVVXPLQJ DV QXOO K\SRWKHV
VDPH SHUIRUPDQFH ,Q WKLV FDVH $F = 19.05$ $p = 1.51e^{-10}$ WKHUH DUH VLJQLILFDQW GLIIHUHQFHV D
ZLWK D ODUJH VL]H HIIHFW 0.29 (WD 6TXDUHG WKH (QVHPEOH GLVWLOODWLRQ PHWKRG SURGXFH
PRGHOV

7HFKQLTXH	'LFH VFRUH	)ULHGPDQ¶V WHVW
'LVWLOODWLRQ		
(QVHPEOH GLVWLOODWLRQ		
3VHXGRODEHO		
&ORVH 7UDQVIHU		
%DVHOLQH		

7DEOH)ULHGPDQ¶V WHVW IRU WKH 'LFH VFRUH RI WKH VHJ

,Q 7DEOH ZH VKRZ WKH UHVXOWV RI WKH DSSOLFDWLRQ RI WKH
FRPSDUH WKH FRQWURO WUDLQLQJ SURFHGXUH ZLQQHU EDVHG
WUDLQLQJ DSSURDFKHV DGMXVWLQJ WKH S YDOXH 5HVXOWV VX
WKH VHPL VXSHUYLVHG OHDUQLQJ SURFHGXUHV DQG WKH SOD
WKH VHPL VXSHUYLVHG OHDUQLQJ PHWKRGV SURGXFH VLPLODU
WDNHQ LQWR DFFRXQW XVLQJ &RKHQ¶V G DQG DV VKRZQ LQ 7DE
WKH ZLQQLQJ DSSURDFK LV FRPSDUHG ZLWK WKH SODLQ WUDLQL

7HFKQLTXH	p YDOXH	p YDOXH	&RKHQ¶V G
%DVHOLQH	H	H	
&ORVH 7UDQVIHU			
3VHXGRODEHO			
(QVHPEOH GLVWLOODWLRQ			

7DEOH $GMXVWHG S YDOXHV ZLWK +ROP DQG &RKHQ¶V G &RQ

&RQFOXVLRQ DQG IXUWKHU ZRUN

,Q WKLV ZRUN ZH KDYH SUHVHQWHG DQ DSSURDFK WR LGHQWL
UHF\FOLQJ SODQW IURP RUWKRSKRWR LPDJHV 2XU PHWKRG ZDV E

OHDUQLQJ PRGHOV IRU VHJPHQWLQJ WKRVH LUUHOHYDQW REM
PXOWL 'LFH VFRUH RI DSSUR[LPDWHOO\ 7KH PDLQ OLPLWDW
ZDV WKH UHGXFHG QXPEHU RI DYDLODEOH DQQRWDWHG RUWRSS
WDFNOHG XVLQJ WZR GLIIHUHQW DSSURDFKHV D FORVH WUDQV
VHPL VXSHUYLVHG OHDUQLQJ WHFKQLTXHV 7KH IRUPHU XVHG
SUH WUDLQHG RQ UDZ LPDJHV WKDW DUH HDVLHU WR DQQRWDWH
WKDW SUH WUDLQLQJ VWDJH WKH SHUIRUPDQFH RI RXU PRGHO L
WKH DSSOLFDWLRQ RI VHPL VXSHUYLVHG OHDUQLQJ PHWKRGV IX
D UDQJH EHWZHHQ DQG RQ DYHUDJH 7KHUHIRUH ZH K
PHWKRGV WR LPSURYH VHJPHQWDWLRQ DFFXUDF\ E\ WDNLQJ D
XQODEHOOHG RUWKRSKRWR LPDJHV WKXV DYRLGLQJ WKH QHH
LPDJHV ZKRVH DQQRWDWLRQ FDQ EH WLPH GHPDQGLQJ

)RU IXUWKHU ZRUN ZH SODQ WR XVH WKH GHYHORSHG PRGHO
REMHFWV IURP WKH LPDJHV DQG WKHQ XVH WKRVH LPDJHV WR
ZHLJKW DW WKH EHJLQQLQJ RI WKH WUHDWPHQW SURFHVV LQ ZD

&RPSOLDQFH ZLWK (WKLFDO 6WDQGDUGV

&RQIOLFW RI ,QWHUHVW $OO WKH DXWKRUV GHFODUH WKDW WK
(WKLFDO DSSURYDO 7KLV DUWLFOH GRHV QRW FRQWDLQ DQ\ VW
DQLPDOV SHUIRUPHG E\ DQ\ RI WKH DXWKRUV

$FNQRZOHGJHPHQWV

7KLV ZRUN ZDV SDUWLDOO\ VXSSRUWHG E\ 0LQLVWHULR GH &LH
5% , $(, @ DQG WKH 6(3$5$ SURMHFW IXQ
&'7, SURMHFW RI WKH 0LVLRQHV FDOO

5HIHUHQFHV

>%DVKNLURYD%HW DO@ %DVKNLURYD ' HW DO =HURZDVWH GDWDVHW 7R
REMHFW VHJPHQWDWLRQ LQ FOXWWHUHG VFHQHV DU;LY SUHSULQW D

>%DVKNLURYD%HW DO@ %DVKNLURYD ' HW DO 9LVXDO GRPDLQ DGDSWDWLR
KWWSV DL EX HGX YLVGD

>%FKLU HW DO@ %FKLU 2 $OJKDQQDP 6 $OVDGKDQ 1 $OVXPDLU\ 5 $
$OPRWOD 0 &RPSXWHU YLVLRQ EDVHG SRO\HWK\OHQH WHUHSK
UHF\FOLQJ ,QWHUQDWLRQDO -RXUQDO RI $GYDQFHG &RPSXWHU 6FLHQFH DQ

>%XFLOD HW DO@ %XFLOD & &DUXDQD 5 DQG 1LFXOHVFX 0L]LO $ 0R
PDNLQJ ELJ VORZ PRGHOV SUDFWLFDO ,Q 3URFHHGLQJV RI WKH WK ,QWHUQDWLRQDO &R
.QRZOHGJH 'LVFRYHU\ DQG 'DWD 0LQLQJ

>&RKHQ &RKHQ - 6WDWLVWLFDO 3RZHU $QDO\VLV IRU WKH %HKDYLR
GHPLF 3UHVV 86$

>&RKHQ &RKHQ - (WD VTXDUHG DQG SDUWLDO HWD VTXDUHG LQ IL
(GXFDWLRQDO DQG 3V\FKRORJLFDO 0HDVXUHPHQW

>'RVRYLWVNL\ HW DO @ 'RVRYLWVNL\ $ HW DO $Q LPDJH LV ZRUWK [ZR
IRUPHUV IRU LPDJH UHFRJQLWLRQ DW VFDOH ,Q ,QWHUQDWLRQDO &RQIHUHQFH RQ /HDUQLQJ 5
WLRQV ,&/5

>*DUFLD HW DO @ *DUFLD HW DO $GYDQFHG QRQSDUDPHWULF WHVWV
LVRQV LQ WKH GHVLJQ RI H[SHULPHQWV LQ FRPSXWDWLRQDO LQWHOOL
DQDO\VLV RI SRZHU ,QIRUPDWLRQ 6FLHQFHV

>*RQ]DOH] HW DO @ *RQ]DOH] 5 & :RRGV 5 ' 'LJLWDO LPDJH SURFHVVLQJ
3UHQWLFH KDOO 8SSHU 6DGGOH 5LYHU 1-

>+LQWRQ HW DO @ +LQWRQ * 9LQ\DOV 2 DQG 'HDQ - 'LVWLOOLQJ WKH
1HXUDO 1HWZRUN

>+RP @ +ROP 6 $ VLPSOH VHTXHQWLDOO\ UHMHFWLYH PXOWLSOH
GLQDYLDQ -RXUQDO RI 6WDWLVWLFV

>+RZDUG DQG *XJJHU @ +RZDUG - DQG *XJJHU 6)DVWDL $ OD\HUHG
OHDUQLQJ IRUPDWLRQ

>+XDQJ HW DO @ +XDQJ / +H - ;X = DQG +XDQJ * $ FRPELQDWL
EDVHG RQ WUDQVIHU OHDUQLQJ &RQFXUUHQF\ DQG &RPSXWDWLRQ 3UDFWL
([SHULHQFH H

>.D]D HW DO @ .D]D 6 < DR / & %KDGD 7DWD 3 DQG :RUGHQ) 9
:DVWH $ *OREDO 6QDSVKRW RI 6ROLG :DVWH 0DQDJHPHQW WR '&
86$

>/H&XQ HW DO @ /H&XQ < %HQJLR < DQG +LQWRQ * 'HHS OHDUQL
±

>/HH @ /HH ' + 3VHXGR ODEHO 7KH VLPSOH DQG HIILFLHQW VHPL
PHWKRG IRU GHHS QHXUDO QHWZRUNV ,&0/ :RUNVKRS &KDOOHQJHV LQ 5HSU
/HDUQLQJ :5(3/

>/HYHQH @ /HYHQH + &RQWULEXWLRQV WR 3UREDELOLW\ DQG 6WDWLVWL
RI +DUROG +RWHOOLQJ 5REXVW WHVWV IRU HTXDOLW\ RI YDULDQFHV SDJ
WR 3UREDELOLW\ DQG 6WDWLVWLFV (VVD\V LQ +RQRU RI +DUROG +RWHO

>/L HW DO @ /L * =DQ - +H 6 /X 4 DQG 0D % 6HPL VXSHUYLVHG
PHQWDWLRQ XVLQJ DGYHUVDULDO OHDUQLQJ IRU SDYHPHQW FUDFN GHW

>/L HW DO @ /L + ;LRQJ 3 $Q - DQG :DQJ / 3\UDPLG DWWHQWLR
VHPDQWLF VHJPHQWDWLRQ DU;LY

>/L HW DO @ /L 5 =KHQJ 6 =KDQJ & 'XDQ & 6X - :DQJ / DQG $WN
0XOWLDWWHQWLRQ QHWZRUN IRU VHPDQWLF VHJPHQWDWLRQ
LPDJHV ,(((7UDQVDFWLRQV RQ *HRVFLHQFH DQG 5HPRWH 6HQVLQJ

>/LQ HW DO @ /LQ 7 < 0DLUH 0 %HORQJLH 6 +D\V - 3HURQD 3 5DPD
3 DQG =LWQLFN & / 0LFURVRIW FRFR &RPPRQ REMHFWV LQ FRQW
RQ FRPSXWHU YLVLRQ 6SULQJHU

>/X HW DO @ /X : &KHQ - DQG ;XH) 8VLQJ FRPSXWHU YLVLRQ W
FRPSRVLWLRQ RI FRQVWUXFWLRQ ZDVWH PL[WXUHV 5HVRXUFHV &RQVHUYDWLRQ DQG 5HF\FOLQJ

>0HQJ DQG &KX @ 0HQJ 6 DQG &KX : 7 $ VWXG\ RI JDUEDJH FODVVLIL
FRQYROXWLRQDO QHXUDO QHWZRUNV ,Q 2QOLQH WUDQVDFWLRQV 7DLZDQ $QDO\WLFV DQG 1HWZRUNV 6\QHUJLHV 7DLZDQ ,(((,&$1

>0HQVLQN HW DO @ 0HQVLQN 7 8LMOLQJV - .X]QHWVRYD $ *\JOL 0 DQG)
)DFWRUV RI LQIOXHQFH IRU WUDQVIHU OHDUQLQJ DFURVV GLYHUVH DSSH
SUHSULQW DU;LY

>2SHQ'URQH0DS @ 2SHQ'URQH0DS 3URMHFW 3\RGP D OLEUDU\ I
FUHDWLQJ RUWKRSKRWRV KWWSV S\RGP UHDGWKHGRFV LR HQ ODW

>2SLW]DQG%XUVW] 2SLW] - DQG %XUVW 6 ( ) 3RSXODU HQVHPEOH PHWKRGV $Q HPSLULFDO VWXG\ DU;LY

>2UWHQ]L HW DO2UWHQ]L@ 9LROLQR 6 3DOORWWLQR))LJRULOOL 6 $QWRQXFFL) ,PSHUL * DQG &RVWD & (DUO\ HVWLPDWLRQ RI GURQH RUWKRSKRWR WKURXJK WKH FDQRS\ UDGLXV

>3DOO\DQG6DPDGGDU3DOO\ 5 @DQG 6DPDGGDU 6 $SSOLFDWLRQ RI LPDJH SURFHVVLQJ DQG FRQYROXWLRQDO QHXUDO QHWZRUNV IRU IORRG LPDJH FODVVL (QYLURQPHQWDO 0RGHOOLQJ 6RIWZDUH

>3DV]NH HW DO3DV]NH@$ HW DO 3\WRUFK $Q LPSHUDWLYH VW\OH GHHS OHDUQLQJ OLEUDU\ ,Q :DOODFK + /DURFKHOOH + %H\JHO]LP DQG *DUQHWW 5 HGLWRUV $GYDQFHV LQ 1HXUDO ,QIRUPDWLRQ 3URFHVVLQJ 6\VWHPV &XUUDQ $VVRFLDWHV ,QF

>3URHQoD DQG 6LPRHV3URHQoD @) DQG 6LP}HV 3 7DFR 7UDVK DQQRWDWLRQV LQ FRQWH[W IRU OLWWHU GHWHFWLRQ DU;LY

>5D]DYLDQ HW DO5D]DYLDQ@$ 6 $]L]SRXU + 6XOOLYDQ - HW DO &11 IHDWXUHV RII WKH VKHOI $Q DVWRXQGLQJ EDVHOLQH IRU UHFRJQLWLRQ ,Q &935:

>6KDSLURQ DQG 6KDSLURQ6KDSLURQ@6 6 DQG :LON 0 % $Q DQDO\VLV RI YDULDQFH WHVW IRU QRUPDOLW\ FRPSOHWH VDPSOHV %LRPHWULND

>6KHVNLQ6KHVNLQ@ ' +DQGERRN RI 3DUDPHWULF DQG 1RQSDUDPHWULF 6WDWLVWLFDO 3URFHGXUHV &5& 3UHVV /RQGRQ

>6LHYHUW6LHYHUW@5 ,QVWDQFH VHJPHQWDWLRQ RI PXOWLFODVV OLWWHU DQG LPEDODQFHG GDWDVHW KDQGOLQJ $ GHHS OHDUQLQJ PRGHO FRPSDULVRQ

>6RXVD HW DO6RXVD@- 5HEHOR $ DQG &DUGRVR - 6 $XWRPDWLRQ RI ZDVWH VRUWLQJ ZLWK GHHS OHDUQLQJ ,Q :RUNVKRS GH 9LVmR &RPSXWDFLRQDO :9& ,(((

>9RJLDW]LV HW DO9RJLDW]LV@$ &KDONLDGDNLV * 0RLURJLRUJRX . /LYDGDV * JLRUJDNL 0 DQG =HUYDNLV 0 'XDO EUDQFK FQQ IRU WKH LGHQWLILFDWLRQ RI UHF\FODEOH PDWHULDOV ,(((,QWHUQDWLRQDO &RQIHUHQFH RQ ,PDJLQJ 6\VWHPV DQG 7HFKQLTXHV ,67 SDJHV ± ,(((

>:DGD HW DO:DGD@ HW DO /DEHOPH LPDJH SRO\JRQDO DQQRWDWLRQ ZLWK S\WKRQ KWWSV JLWKXE FRP ZNHQWDUR ODEHOPH

><DQJ HW DO<DQJ@ 6 : 3DUN + - .LP - 6 &KRL : 3DUN - DQG +DQ 6 6WXG\ RQ WKH UHDO WLPH REMHFW GHWHFWLRQ DSSURDFK IRU OLWWHU UHF\FOLQJ $SSOLFDEOH DW 6651

>=KRX HW DO=KRX =@ 5DKPDQ 6LGGLTXHH 0 0 7DMEDNKVK 1 DQG /LDQJ - 8QHW $ QHVWHG X QHW DUFKLWHFWXUH IRU PHGLFDO LPDJH VHJPHQWDWLRQ ,Q 'HHS OHDUQLQJ LQ PHGLFDO LPDJH DQDO\VLV DQG PXOWLPRGDO OHDUQLQJ IRU FOLQLFDO GHFLVLRQ

>=KX DQG *ROGEHUJ=KX; DQG *ROGEHUJ $ % ,QWURGXFWLRQ WR VHPL VXSHUYLVHG OHDUQLQJ 6\QWKHVLV OHFWXUHV RQ DUWLILFLDO LQWHOOLJHQFH DQG PDFKLQH