DEEP LEARNING FOR CODED TARGET DETECTION

Vladimir V. Kniaz^{a,b}, Lev Grodzitskiy^a, Vladimir Knyaz^{a,b}

^a State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia (vl.kniaz, lev.grodzickiy)@gosniias.ru
^b Moscow Institute of Physics and Technology (MIPT), Russia

Commission II, WG II/8

KEY WORDS: photogrammetry, deep learning, coded target detection

ABSTRACT:

Coded targets are physical optical markers that can be easily identified in an image. Their detection is a critical step in the process of camera calibration. A wide range of coded targets was developed to date. The targets differ in their decoding algorithms. The main limitation of the existing methods is low robustness to new backgrounds and illumination conditions. Modern deep learning recognition-based algorithms demonstrate exciting progress in object detection performance in low-light conditions or new environments. This paper is focused on the development of a new deep convolutional network for automatic detection and recognition of the coded targets and sub-pixel estimation of their centers.

1. INTRODUCTION

Coded targets are physical optical markers that can be easily identified in an image. Their detection is a critical step in the process of camera calibration. Usually, each target has its unique ID that can be decoded from its shape or color. A wide range of coded targets was developed to date. The targets differ in their decoding algorithms.

Most of the algorithms include three steps for target recognition. Firstly, target locations are detected in the image. After that, target IDs are decoded to matching their locations in image and object coordinate systems. Finally, target area is analyzed for estimation of sub-pixel coordinates of the target's center. The main limitation of the existing methods is low robustness to new backgrounds and illumination conditions.



Figure 1. Example results of coded targets detection by our VOLO-Target model.

Modern deep learning recognition-based algorithms demonstrate

exciting progress in object detection performance in low-light conditions or new environments. Still, deep-learning-aided coded target recognition received little scholar attention to date. This paper is focused on the development of a new deep convolutional network for automatic detection and recognition of the coded targets and sub-pixel estimation of their centers.

2. RELATED WORK

Coded targets is important tool for automation of point identification and stereo matching, especially for photogrammetric applications addressing for weak-textured objects. They allow to solve automatically the correspondence problem in calibration and orientation tasks. Therefore the design of coded targets that and provide the high recognition and measurement performance, attracts attention of photogrammetric and computer vision researchers.

The first studies addressed to the development of targets that can be automatically detected in an image has been fulfilled at the end of XX century (Russo, Knockeart, 1972), (Fraser, 1997), (Knyaz, Sibiryakov, 1998). Since then many different configuration of coded targets and techniques for targets detection and identification has been proposed. The efforts of researches were aimed at designing more robust targets and accurately detecting their image coordinate.

First hand-crafted algorithms used dots, segments or colours to encode target's identification (ID).

A coded target system (Shortis et al., 2003) consists of the central circular target and a surrounding square. The detecting algorithm is based on Hough transform, and segment matching is used for automatic recognition and identification of the targets in digital images. The algorithm exploits a pre-detection processes, developed to improve the performance under unfavourable conditions. The improved coded target system (Shortis, Seager, 2014) was widely used in a variety of calibration and measurement applications.

Dot-distributed coded targets (Feng et al., 2010) for digital industrial photogrammetry was composed of 5 template points and 3 code points. Each coded target contained different character strings as a coder, max code capacity of the target being 496. The decoding algorithm was based on affine transform and cross ratio analysis, and allowed to identify the targets automatically and rapidly. The applicability of this type of dot-distribution coded targets for the task of the automatic digital industrial photogrammetry has been proved experimentally.

The color coded target for vision measurements (Yang et al., 2014) uses a pair of concentric circles for the precise location of the target. For avoiding repeated coding and for simplifing the structure of the code a reference position is defined. The edge-based identification and location algorithm exploits the coarse-to-fine strategy for improving the efficiency and the correctness of target identification. The proposed algorithm demonstrated high accuracy and robustness in experiments performed.

Analysis of grey gradient for coded target detection (Ran et al., 2016) is utilized to estimate the central angles of each coded section. The algorithm for accurate ellipse detection extracts ellipse centres at sub-pixel level, and eliminates false ellipses with great best-fit error. Gradient analysis of the coded band is used obtain the central angle of each coded section.

The combined circular dot-distribution coded target were proposed to improve the coding capacity and accuracy of coded target identification in large-scale industrial photogrammetry (Jingui, Liyuan, 2020). These coded targets were used for point matching and image orientation in large-scale photogrammetry applications, and also were applied for the automatic management of a large number of measuring points.

Influence of target parameters on their performance was studied as from the point of view target shape, so from the point of view target design and size. A practical method for selecting a proper size of circular targets (Liu et al., 2019), basing on the theoretical model of perspective imaging. The authors experimentally demonstrated that the measured semi-major axis of the image ellipse corresponding to the circular target agreed well with its theoretical design value. The maximum difference is estimated at the level of 0.5 pixel, that, in fact, is not accepable for the practice of accurate photogrammetric measurements.

To overcome coded low performance of coded targets in degraded images, such as motion blur images, Chinese character coded targets has been proposed (Shi, Zhang, 2020). These targets are of square form with a relatively small circular feature overlaid in the middle of a square Chinese character. To facilitate extraction of the center point of a target, central black circle contains white circular ring, serving as a circular feature for target localization. The Chinese character is the square out region serves for target identification. The Faster Region-based Convolutional Neural Network (Ren et al., 2017) is trained for recognition of the targets in motion blurred images.

The concentric circular coded targets designed for applying in complicated conditions of poor illumination and flat viewing angles (Yan et al., 2021) utilize practical error-compensation model to correct the eccentricity errors. To improve recognition abilities in poor illumination, such as overexposure and underexposure, an adaptive brightness adjustment has been proposed. Also four vertices of the background area of the target added to improve the robustness of the recognition by perspective correction.

3. METHOD

The coded targets used in the current study has been developed for automatic target detection and identification and sub-pixel image coordinates measurements (Knyaz, 2006). They were successfully applied in various applications, such as motion capture (Knyaz, 2015), pose estimation for unmanned aerial vehicle (Kniaz, 2016), photogrammetric system calibrations (Knyaz, Zheltov, 2017, Knyaz, 2021) etc., providing high robustness of target recognition and high accuracy of target center measurement.

3.1 Hand-crafted algorithm for target recognition

Current algorithm for target recognition firstly detect all image regions meeting the given requirements and then approximate the border of each detected region by ellipse and extract target code. The full procedure includes the following steps (Algorithm 1):

Algorithm 1: Coded target recognition				
Input:				
Image $\mathbf{I} \in \mathbb{R}^{h \times w \times 3}$,				
Parameters of coded target $\mathbf{P} = \{p^k\}$				
Output:				
Array of coordinates and codes $\{x_0^t, y_0^t, ID^t\}$ for detected				
targets in the image I				
Target detection;				
Procedure Target detection(<i>I</i>):				
Binarize the image <i>I</i> ;				
Detect eight-connected regions R^8 not laying in the image				
border $G(I)$: $\mathbb{R}^8 \wedge \{G(I)\};$				
for each R_i^8 do				
if R_i^8 being a structure "black region inside white				
region" then				
if Parameters of of the structure $\hat{R}_i^8 \in \{p^k\}$ then				
save \hat{R}_{i}^{8} in possible coded targets array R_{n}^{t} ;				
else				
skip;				
end				
else				
skip;				
end				
end				
return $\mathcal{R}^t = \{r_j^t\};$				
Code Reading ;				
Procedure Code Reading($\mathcal{R}^t = \{r_i^t\}$):				

for each target region from the array R_i^t do				
Approximate internal and external border by ellipse \mathcal{E}_j ;				
Determine by least squares minimisation the parameters				
of the ellipse \mathcal{E}_j :				
subpixel centre coordinates x_0, y_0 ,				
sizes of a large and small half-axis d_l, d_s ,				
inclination of a large half-axis β ;				
Transform coding circle to rectangle;				
Extract target code <i>ID</i> ;				
end				
return $\{x_0^t, y_0^t, ID^t\};$				

3.1.1 Binarization of the image. Binarization of the image is used for detecting the borders of the targets. A set of thresholds is applied for image binarization for compensate possible differences in local intensity and contrast. The result of the thresholding is a set of binary images B_i . Then searching for targets

is performed at each binary image. Binarization of the image is carried out by thresholding:

$$B_i(x, y) = 0, f(x, y) < t_i(x, y) (1)$$

$$B_i(x, y) = 255, f(x, y) > t_i(x, y) (2)$$

where f(x, y) - original image, b(x, y) - binary image, $t_i(x, y)$ - threshold level.

Constant threshold $t(x, y) = t_0$ applying has significant drawback caused by non-uniform contrast for different parts of the image, and as consequence, possibility of loosing targets during binarization. Therefore a set of the thresholds is applied, with further tracking the detected targets through the set of binary images B_i .

All eight-connected regions R^8 are detected In each binary image B_i , excluding those of them that lay on the border of the image. Then each detected region R^8 is checked to be the structure "white region that is surrounded by black region, that is surrounded by white region, that is surrounded by black region" $R_{w \in b \in w \in b}$ (Figure 2).



Figure 2. Coded target structure: Outer border serves for elliptic approximation of the target; Coding strip stores the ID of the target; "Anchor" define the initial position for code reading; Centre can be used for correction of target coordinates.

All detected regions $R_{w \in b \in w \in b}$ are verified to be a region having correct geometric parameters, such as ratio of areas of bounding boxes of targets' elements (Figure 2).

3.1.2 Elliptic approximation The approximation of the target by an ellipse solves two tasks:

- 1. Identifying target region as a projection of a circle;
- 2. Measuring the position of the target with sub-pixel accuracy.

The border of the target region should be described by the equation of an ellipse:

$$Ax^{2} + Bxy + Cy^{2} + Dx + Ey + F = 0$$
(3)

To solve the problem of the ellipse parameters estimation by least mean squares adjustment, it is more convenient to write down the equation of an ellipse in the following form:

$$ax^2 + bxy + cy^2 + dx + ey = -1 + \epsilon \tag{4}$$

Or in matrix notation Equation 4 can be written as:

$$A \cdot \boldsymbol{p} = \boldsymbol{q} + \boldsymbol{\epsilon} \tag{5}$$

where $\boldsymbol{p} = (a, b, c, d, e)^T$ - a vector of equation parameters; A - design matrix with *i*-th line corresponding to *i*-th pixel of the region border: $[x/x; y; y; 2x; y;]; \boldsymbol{q} = [-1, ..., -1]^T$ - vector of observations; ϵ - error vector.

The solution of the Equation 5 is:

$$\boldsymbol{p} = (\boldsymbol{A}^T \cdot \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{q} \tag{6}$$

The parameters of the ellipse, needed for target location measurement, are given by the following expressions:

• center of the ellipse:

$$x_0 = \frac{be - 2cd}{4ac - b^2} \tag{7}$$

$$y_0 = \frac{bd - 2ae}{4ac - b^2} \tag{8}$$

• inclination of a large semi-axis β

$$\beta = \frac{1}{2}atan(\frac{b}{a-c}) \tag{9}$$

large semi-axis d_l

$$d_{l} = \frac{1}{\sqrt{a \cdot \cos^{2}(\beta) + b \cdot \sin\beta \cos(\beta) + c \cdot \sin^{2}(\beta)}}$$
(10)

small semi-axis d_s

$$d_s = \frac{1}{\sqrt{c \cdot \cos^2(\beta) - b \cdot \sin\beta \cos(\beta) + a \cdot \sin^2(\beta)}}$$
(11)

The ellipse should satisfy following invariants

$$\frac{I_3}{I_1} \le 0 \tag{12}$$

$$I_2 \ge 0 \tag{13}$$

where I_1 , I_1 , I_3 are:

$$I_1 = a + c \tag{14}$$

$$I_2 = \begin{vmatrix} a & b/2 \\ b/2 & c \end{vmatrix} \tag{15}$$

$$I_3 = \begin{vmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & 1 \end{vmatrix}$$
(16)



Figure 3. YOLO-Target model. It divides the image an S x S grid for each grid cell predicts B bounding boxes with centers, confidence for those Boxes, and C class probabilities. These predictions are encoded as an S x S x (B * 7 + C) tensor.

The invariants I_1 , I_1 , I_3 are used for verifying if the detected region is an ellipse.

3.1.3 Target code reading. For each detected region, that passed the "elliptic" check the coding strip is extracted, and the target *ID* is retrieved. To extract the coding strip

The following coordinate transform:

$$\xi = x_0 + x \cdot \cos(\beta) + y \cdot \sin(\beta) \tag{17}$$

$$\eta = y_0 - x \cdot \sin(\beta) + y \cdot \cos(\beta) \tag{18}$$

is applied to extract the coding strip. Then parametrical equations of an ellipse can be written as the following:

$$\xi = d_l \cdot \cos(t); \ \eta = d_s \cdot \sin(t) \tag{19}$$

where t is a parameter that varies from 0 to 2π .

The extracted area of the coding strip is then transformed to rectangular one (Figure 4).

If $x_0^1, y_0^1, \beta^1, d_l^1, d_s^1$ are the parameters of the internal ellipse, and $x_0^2, y_0^2, \beta^2, d_l^2, d_s^2$ are the parameters of an external ellipse, then eectangular transform of a coding strip K(d, t) using original image f(x, y) and parametrical equations of ellipse (19) is given by:

$$K(d,t) = f(x,y) \tag{20}$$

$$\xi = d \cdot \cos(t); \ \eta = d \cdot \sin(t) \tag{21}$$

$$x = x_0^1 + \xi \cdot \cos(\beta^1) - \eta \cdot \sin(\beta^1) \tag{22}$$

$$y = y_0^1 + \xi \cdot \sin(\beta^1) + \eta \cdot \cos(\beta^1) \tag{23}$$

$$t \in [0, 2\pi] \tag{24}$$

The example of code reading for a target is shown in Figure 4. Coding strip is extracted using the target parameters. Inside the strip (Figure 4) grey-level profile is binarized by Otsu criterion (Otsu, 1979).



(a) Image of the coded target



(b) Rectangular area of the coding strip

Figure 4. Example results of code reading.

3.2 YOLO-Target Convolutional Network

Our YOLO-TARGET is inspired by the YOLOv3 model. It solves two tasks. Firstly, it detects the bounding box of the coded target. Secondly, it estimates the position of the target center. The details of our YOLO-TARGET architecture are presented in Table 1 and Figure 3.

3.3 Dataset Generation

To train our YOLO-TARGET framework, we collected a new dataset CodeSet. It includes synthetic images of calibration page with 49 coded marks (see figure 1). To increase the variability of the training dataset, we used Blender3D creation suite. It is a free and open-source 3D computer graphics software tool set used for

	Туре	Filters	Size	Output
	Convolutional	32	3×3	256×256
	Convolutional	64	3 imes 3 / 2	128×128
	Convolutional	32	1×1	
$1 \times$	Convolutional	64	3×3	
	Residual			128×128
	Convolutional	128	3 imes 3 / 2	64×64
	Convolutional	64	1×1	
$2 \times$	Convolutional	128	3×3	
	Residual			64×64
	Convolutional	256	3 imes 3 / 2	32×32
	Convolutional	128	1×1	
$8 \times$	Convolutional	256	3×3	
	Residual			32×32
	Convolutional	512	$3 \times 3 / 2$	16×16
	Convolutional	256	1×1	
$8 \times$	Convolutional	512	3×3	
	Residual			16×16
	Convolutional	1024	3 imes 3 / 2	8×8
	Convolutional	512	1×1	
$4 \times$	Convolutional	1024	3×3	
	Residual			8×8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Table 1. The YOLO-TARGET architecture.

creating animated films, visual effects, art, 3D printed models, motion graphics, interactive 3D applications. We exported the calibration page texture image to 3D creator and developed special software using Python3 that allows you to randomly position and rotation the calibration page on a frame. The calibration page was shot from different angles. We used different turning HDR textures as background. A part of the dataset was created with low light conditions and with adding "salt and pepper" noise effect. Label images were geometric aligned with color images. Each coded mark had has a unique RGB-code that allowed to mark up the generated data in Y0L0 format and transferred it to the neural network for training. We synthesized about 1000 images for network training and about 100 images for evaluation results.



Figure 5. Examples from the dataset.

4. EXPERIMENTS

The algorithm was evaluated using the generated dataset. The following section presents details on the learning of the YOLO-TARGET and the perceptual and the quantitative evaluation of our algorithm.

4.1 Network Training

Training was carried out on a video card Nvidia Geforce GTX 1080 with 8GB video memory for 4 hours. The result of training a neural network is shown in the figure 1. As you can see from the figure, coded labels are well recognized and identified.

4.2 Qualitative Evaluation

We evaluate our model and baseline qualitatively by running inference on challenging images (Figure 6). Qualitative evaluation proves that our YOLO-Target model successfully detects targets in challenging configurations and low-light conditions.

Original	EO	YOLO-Target

Figure 6. Detection results.

4.3 Quantitative Evaluation

The results of the evaluation are encouraging and demonstrate that our YOLO-Target model outperforms baselines by a large margin. Specifically, we evaluate our model and baselines in terms of target localization intersection over union, in mAP of target ID recognition, and endpoint error of the target center estimation with sub-pixel accuracy. Our model increases the target recognition mAP by 40% compared to the loading baselines. The endpoint error estimation accuracy is comparable or better than the accuracy of baselines.

5. CONCLUSION

We developed an end-to-end pipeline for localization and recognition of circular coded targets using a deep convolutional network. Moreover, our model is capable of estimating the target's center with sub-pixel accuracy. Our model provides real-time performance for 720p input video on modern GPUs. We made our code and dataset publicly available.

	mAP	IoU	MSE T. Center
YOLO-TARGET@1280	0.61	0.66	0.23
YOLO-TARGET@600	0.65	0.68	1.2
TargetDetector	0.72	0.83	0.24

Table 2. Results of comparison new and old methods.

ACKNOWLEDGEMENTS

The reported study was funded by Russian Foundation for Basic Research (RFBR) according to the research project 19-29-13040.

REFERENCES

Feng, Q., Zongchun, L., Guangyun, L., 2010. Design and decoding of dot-distribution coded targets. *8th World Congress on Intelligent Control and Automation.* https://www.semanticscholar.org/paper/Designand-decoding-of-dot-distribution-coded-Feng-Li/535114e9b9cffb41e30524a4d9d98fb1e8e1c823.

Fraser, C. S., 1997. Innovations in Automation for Vision Metrology Systems. *The Photogrammetric Record*, 15(90), 901-911. https://onlinelibrary.wiley.com/doi/abs/10.1111/0031-868X.00099.

Jingui, Z., Liyuan, M., 2020. Design of a New Coded Target With Large Coding Capacity for CloseRange Photogrammetry and Research on Recognition Algorithm. *IEEE Access*, 8. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9285236.

Kniaz, V., 2016. ROBUST VISION-BASED POSE ESTIMA-TION ALGORITHM FOR AN UAV WITH KNOWN GRAVITY VECTOR. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 63-68.

Knyaz, V. A., 2006. Automated calibration technique for photogrammetric system based on a multimedia projector and a CCD camera. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVI-5. https://www.isprs.org/proceedings/XXXVI/part5/paper/ 1230_Dresden06.pdf.

Knyaz, V. A., 2015. SCALABLE PHOTOGRAMMETRIC MO-TION CAPTURE SYSTEM "MOSCA": DEVELOPMENT AND APPLICATION. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W6, 43–49. https://www.int-arch-photogrammremote-sens-spatial-inf-sci.net/XL-5-W6/43/2015/.

Knyaz, V. A., 2021. Optical system calibration for 3d measurements in a hydrodynamic tunnel. *Computer Optics*, 45, 58–65.

Knyaz, V. A., Sibiryakov, A., 1998. The Development of New Coded Targets for Automated Point Identification and Non-Contact 3D Surface Measurements. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXII-B5, 80–85. http://www.isprs.org/proceedings/XXXII/part5/80_XXXIIpart5.pdf. Knyaz, V., Zheltov, S., 2017. Accuracy evaluation of structure from motion surface 3D reconstruction. *Proc.SPIE*, 10332, 10332 - 10332 - 10. https://doi.org/10.1117/12.2272021.

Liu, Y., Su, X., Guo, X., Suo, T., Li, Y., Yu, Q., 2019. A novel method on suitable size selection of artificial circular targets in optical non-contact measurement. *Applied Optics and Photonics China*. https://www.semanticscholar.org/paper/A-novel-method-on-suitable-size-selection-of-in-Liu-Su/7af88bc8ee9729fdf34779554e7a2d4c2c06ba1a.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.

Ran, C., Zhong, K., Li, Z., Meng, L., Guomin, Z., 2016. An accurate and reliable circular coded target detection algorithm for vision measurement. *SPIE/COS Photonics Asia*. https://www.semanticscholar.org/paper/Anaccurate-and-reliable-circular-coded-target-for-Chen-Zhong/137d118b3c7126958287973f3bd814a1c414c410.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149.

Russo, F. A., Knockeart, R. P., 1972. Automated data acquisition with an optical code reader. *Bendix Technical Journal*, 5, 48 - 52.

Y., Zhang, L., 2020. Design of Chinese Char-Shi, acter Coded Targets for Feature Point Recog-Motion-Blur Effect. IEEE nition Under Access, XX. https://www.semanticscholar.org/paper/Designof-Chinese-Character-Coded-Targets-for-Point-Shi-Zhang/e04528fbef2267270bc7ed9b9e1fdded1ae210d8.

Shortis, M. R., Seager, J. W., 2014. A practical target recognition system for close range photogrammetry. *The Photogrammetric Record*, 29(147), 337-355. https://onlinelibrary.wiley.com/doi/abs/10.1111/phor.12070.

Shortis, M. R., Seager, J. W., Robson, S., Harvey, E. S., 2003. Automatic recognition of coded targets based on a Hough transform and segment matching. S. F. El-Hakim, A. Gruen, J. S. Walton (eds), *Videometrics VII*, 5013, International Society for Optics and Photonics, SPIE, 202 – 208.

Yan, L., Xin, S., Xiang, G., Tao, S., Qifeng, Y., 2021. A Novel Concentric Circular Coded Target, and Its Positioning and Identifying Method for Vision Measurement under Challenging Conditions. *Sensors*, 21(3). https://www.mdpi.com/1424-8220/21/3/855/htm.

Yang, X., Fang, S., Kong, B., Li, Y., 2014. Design of a color coded target for vision measurements. *Optik.* https://www.semanticscholar.org/paper/Designof-a-color-coded-target-for-vision-Yang-

Fang/0173eb77b27c961d9af3341ab936d5db9887c4a5.