DUAL-CYCLE DEEP REINFORCEMENT LEARNING FOR STABILIZING FACE TRACKING

*Congcong Zhu, Zhenhua Yu, Suping Wu, Hao Liu**

School of Information Engineering, Ningxia University, Yinchuan, 750021, China

ABSTRACT

In this paper, we propose a dual-cycle deep reinforcement learning (DCDRL) method for stabilizing face tracking. Unlike most existing face tracking approaches which require per-frame annotations and dense facial landmarks are usually quite costly to annotate manually, our DCDRL aims to learn a robust face tracking policy by only using weakly-labeled annotations that were sparsely collected from raw video data. Motivated by the fact that facial landmarks in videos are usually coherent along with the forward and backward playing orders, we formulate the face tracking problem as a dual-cycle Markov decision process (MDP) by defining two agents for the forward-cycle and the backward-cycle accordingly. Specifically, both agents reason with the MDP policies by interacting in tuples of states, state transitions, actions and rewards during the MDP processes. Moreover, we carefully design a consistency-check reward function to track along until the target and back again it should arrive the start position in the reverse order. With the designed function, each policy generates a sequence of actions to refine the tracking routing by accumulating the maximal scalar rewards. This typically enforces the temporal consistency constraint on consecutive frames for reliable tracking outcomes. Experimental results demonstrate the robustness of our DCDRL versus many severe challenging cases especially in uncontrolled conditions.

Index Terms— Face tracking, video-based face alignment, deep reinforcement learning, biometrics.

1. INTRODUCTION

Face tracking, *a.k.a.*, video-based face alignment, aims to localize a serious of multiple facial landmarks for a given face sequence, which plays a vital step for many facial analysis tasks [1,2]. The main challenges for unconstrained face tracking is the stabilization versus types of facial variations due to diverse temporal motions in videos. Moreover, densely annotating each frame for a large scale video consumes much effort. Hence, both issues motivate us to propose a robust face tracking method by using limited and partial annotations.

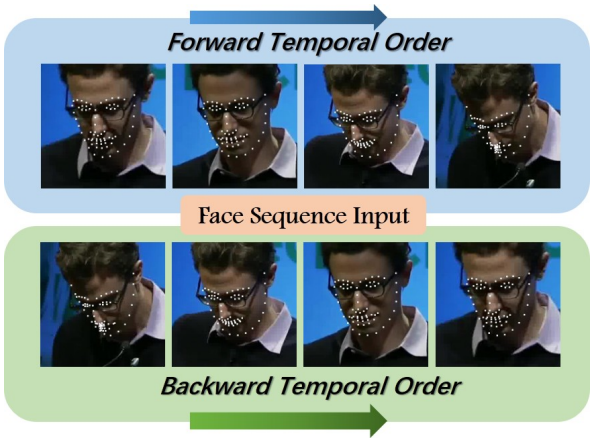

Fig. 1. Insight of our DCDRL. Observing from these ordered frames, we figure out that the bidirectional temporal orders undergo subtle effects to landmark movements across frames. In our approach, we propose a reinforcement learning method to seek a reliable face tracking policy by simultaneously exploiting the dual and complementary information from both orders, *i.e.*, playing the input video forward and backward.

Conventional face tracking methods can be roughly classified into two-fold categories: image-based and video-based. Image-based methods [3–9] intend to seek a sequence of discriminative feature-to-shape mappings, so that the initialized shape is adjusted to the target one in a coarse-to-fine manner. To make the image-based methods adaptive for video data, one common solution is to regard the outcomes of previous frames as initializations for the following frames via a tracking-by-detection method [10]. However, this method could only extract 2D spatial appearances from still images and cannot explicitly exploit temporal information on consecutive frames. To circumvent this problem, video-based methods [11–14] learn to memorize and flow the temporal consistency information across frames, which improves the robustness to the jitter problem in visual tracking. One major issue lies on these methods is that they require a volume of per-frame annotations to train their models, where it is costly to manually label especially for large scale video data.

* Corresponding Author is Hao Liu (e-mail: liuhao@nxu.edu.cn). This work was supported in part by the Natural Science Foundation of Ningxia under Grant 2018AAC03035, in part by the National Natural Science Foundation of China under Grant 61662059 and Grant 61806104, in part by the Scientific Research Projects of Colleges and Universities of Ningxia under Grant NGY2018050, and in part by the Youth Science and Technology Talents Enrollment Projects of Ningxia under Grant TJGC2018028.

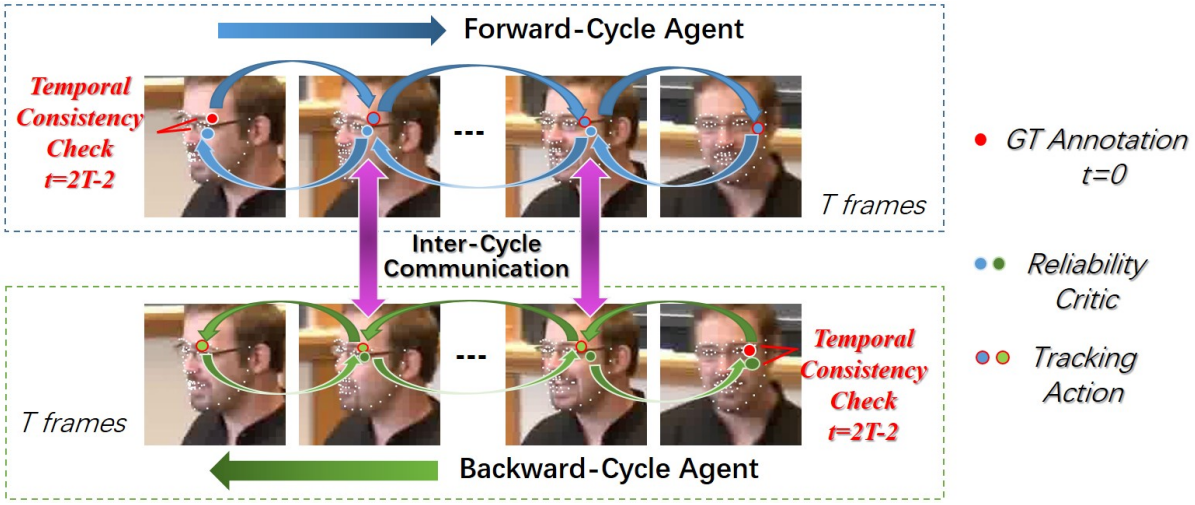

Fig. 2. Illustration of the proposed *dual-cycle* execution in our DCDRL. Specifically, our DCDRL acts in two dual agents where each agent manages the forward-cycle and backward-cycle accordingly. Taking the forward-cycle denoted by blue arrows as an example (we select one landmark for better visualization in this figure), our agent reasons a series of tracking actions until the target frame and back again it should arrive the starting frame in the reverse order. Moreover, we develop a designed temporal consistency check function to efficiently evaluate the tracking reliability. This enforces the temporal consistency constraint on the consecutive frames. During training procedure, both agents are optimized within both cycles in a cooperative and competitive manner. This figure is viewed in color pdf file and under zoom.

Apart from taking full access to completed training labels, self-supervised learning is proposed to predict a set of plausible pseudo-labels by defining a proxy task, so that the supervisions are dominantly augmented without using additional labels. These label-augmentation methods such as [15–17] dramatically enhance robust model training and discriminative representation learning. In the term of pseudo-labels in temporal modeling, Wei *et al.* [16] developed an unsupervised learning method to verify video playing orders (forward and backward), where these extracted axillary cues contribute improvements on action recognition. Meister *et al.* [17] developed an unsupervised learning approach specific for computing robust optical flow by designing a bidirectional census loss in their formulation. However, the aforementioned methods ignore the geometric deformation of facial landmarks due to the 3D-2D projection [18], which cannot be straightforwardly applied for deformable face tracking. To circumvent this, Dong *et al.* [19] introduced a semi-supervised method with the LK tracker [20], which proceeds the dense correspondences of the between-frame landmarks. More specifically, the employed LK-tracker localizes facial landmarks forward and then evaluates the estimated results backward the video. Nevertheless, the performance is restricted because it ignores the intrinsic connections between the bidirectional temporal orders as described in Fig. 1, which provides dual and complementary information for stabilizing face tracking.

To address above-mentioned challenges, in this paper, we propose a dual-cycle deep reinforcement learning (DCDRL) method by using weakly-supervised signals. Unlike existing fully-supervised tracking methods, our DCDRL aims to learn an optimal face tracking policy by computing the cumulative scalar rewards. Motivated by the inspiration that facial landmarks are temporally correlated along the bidirectional-cycle orders (playing forward or backward) in videos, we formulate the problem of tracking in both orders as a dual-cycle Markov decision process (MDP) with two agents, by interacting based on tuples of states, state transitions, actions and rewards. To achieve this, both agents simultaneously reasons with the forward-cycle and backward-cycle policies within the MDP process. Each policy accepts a set of raw patches directly from the facial image as the input. Then it generates a sequence of residuals to make decisions on the plausible tracking routing across frames. To further evaluate the reliability of each cycle, we compute the consistency-check reward to enforce on tracking results which should be retraced back again to the start. In this way, our policy makes reliable tracking results by preserving the temporal consistency constraint. During training procedure, we jointly optimize both policies cooperatively and competitively by following the multi-agent deterministic policy gradient algorithm, providing an inter-cycle message passing for discriminative policy inferences. Fig. 2 specifies our architecture under the dual-cycle MDP process. Experimental results show the effectiveness of the proposed approach on the widely-evaluated video-based face alignment dataset.

The core contributions are summarized as follows:

1) We propose a deep reinforcement learning method to address the stability issue specific for semi-supervised face tracking. With only weakly-supervised signals, our architecture reasons with bidirectional temporal orders playing the raw input video forward and backward simultaneously, so that more dual and complementary information is exploited for reliable tracking results.

2) We carefully define a temporal consistency-check reward function to efficiently evaluate our tracking reliability. With the computed reward, our architecture enforces that the tracking results should arrive the start location in the reverse order, and moreover teaches that our performance degrades to the backbone detector versus the jitter problem due to severe occlusions.

2. DUAL-CYCLE DEEP REINFORCEMENT LEARNING

2.1. Problem Formulation

We let each video clip denoted by $\{(\boldsymbol{I}^t, \boldsymbol{p}^t)\}_{t=0}^{T}$ with T frames, where $\boldsymbol{I}^t$ represents the detected raw face, $\boldsymbol{p}^t = [p_1, p_2, \cdots, p_L] \in \mathcal{P} \in \mathbb{R}^{2\times L}$ denotes the shape vector at the t-th frame, respectively. We let $\boldsymbol{p}^* = [p_1^*, \cdots, p_L^*]$ denote the GT annotations, where those of the starting frame and the ending one are exposed to the training procedure.

State and Action: We let the tracking movements $\boldsymbol{a} \in \mathbb{R}^{2\times L}$ over a continuous space as the MDP *action*, which means an offset to refine the positions of all landmarks onto the following frames. The MDP *state* in our approach is defined by a set of partial observation $\boldsymbol{s}{=}o(\boldsymbol{I}, \boldsymbol{p}) \in \mathcal{S} \in \mathbb{R}^{d\times d\times L}$ (ignoring t for simplicity), which are locally cropped directly from the raw facial image via a widely-utilized shape-indexed manner [3, 4, 6], where d denotes the length of each local patch.

As illustrated in Fig. 2, the state $\boldsymbol{p}^0$ is performed by the backbone detector, *i.e.*, MDM [6] and the desired state requires to semantically parse the whole face via different parts including two eyes, eyebrows, nose, mouth and facial cheek. Moreover, our agents reasons with the bidirectional cycles by playing the raw input video forward and backward. Starting from the 0-th frame, each agent produces $T-1$ actions onto the target frame and then goes back to the start by $T-1$ actions. Hence, our tracking result for the last action terminates at the $2T-2$ time stamp.

State Transitions: For face tracking problem, we define two types of the MDP *state transitions* which incorporate both the appearance transition and the facial landmark transition. The appearance transition aims to capture the variations in facial appearance due to the temporal motions. Respectively, the landmark transition is used to refine the positions of all landmarks by the emitted actions across frames. To clarify this, we take an emitted action $\boldsymbol{a}^t$ at the t-th frame as an example, the shape vector is adjusted by the landmark transition $\boldsymbol{p}^{t+1} = \boldsymbol{p}^t + \boldsymbol{a}^t$. Meanwhile, the partially-observed patches is shifted by the appearance transition as $\boldsymbol{s}^{t+1} = o\left(\boldsymbol{I}, \boldsymbol{p}^{t+1}\right)$.

Consistency-Check Reward: Our *reward* function is designed to measure the misalignment error tracking forward and then it should arrive the starting position in reverse order, which is partially inspired by [21] and defined as follows:

$$r(\boldsymbol{s}^t, \boldsymbol{a}^t) = \begin{cases} 1, & m^t \le \epsilon_1 \quad \text{and} \quad t = 2T-2, \\ 0, & m^t > \epsilon_1 \quad \text{and} \quad t = 2T-2, \\ -1, & \text{Det}(\boldsymbol{I}^t) - \boldsymbol{p}^t > \epsilon_2, \quad 0 < t < 2T-2, \end{cases}$$

where ϵ_1 and ϵ_2 denote two thresholds where we specified $\epsilon_1 = 0.3$ and $\epsilon_2 = 0.5$ in our experiments, Det$(\cdot)$ indicates a backbone detector (we used MDM [6] in the experiments), the $m^T = \frac{\|\boldsymbol{p}_i^T - \boldsymbol{p}_i^*\|}{\zeta}$ at the T-th iteration since the groundtruth will be known to each training sequence back again to the start frame, $\|\cdot\|$ specifies the ℓ_2 norm, and ζ denotes the inter-pupil distance as the normalizing factor [3, 5], receptively.

Our reward function typically justifies two-fold scenarios: 1) The higher reward will be given when the tracking results go back again to the start, which enforces the temporal continuity in the learned policies. 2) A negative feedback is provided when the discrepancy between tracking results across frames and those generated by an image-based detector. This means the performance degrades to the pre-trained backbone detector when we undergo the tracking lost issue. It should be noted that both the forward-cycle agent and backward-cycle agents are well-controlled by these time-delayed reward signals, so that our method encodes the temporal consistency information for reliable and robust face tracking results.

Policy Network: We let π to specify the MDP *policy* over a large continuous shape space, which aims to reason a face tracking routing by accumulating a plausible temporal consistency check rewards. Since making full access to the large scale action space is costly [22] especially during the training process, we directly leverage a deterministic and differential policy function $\boldsymbol{a} = f_\pi(\boldsymbol{s})$, which is represented by using a deep convolutional neural network. Benefiting from the nonlinearity of the deep architecture, our policy network is used to exploit the nonlinear mapping between the pairs of states and actions. In our approach, we have two dual agents where one agent capture the forward-cycle tracking process and the other depicts the dual backward-cycle process. To jointly optimize policy networks of both agents, we applied a multi-agent actor-critic policy gradient [23] to optimize both policies in a cooperative and competitive manner.

Besides from the used policy network, we deploy a critic network denoted by $Q_\pi(\boldsymbol{s}, \boldsymbol{a})$ to evaluate the reliability of the tracking results for the dual-cycle executions, *i.e.*, the forward- and backward-cycle. Specifically, the critic network is fed with the cropped local patches based on the resulting landmarks to predict the confidence score. Hence, we directly use the policy network architecture as the critic network specification and the only revision is to append the last layer

to regress the one-dimension score. In addition, we find out making a copy of our policy network for the network initialization achieves sufficient performance in our task.

Objective Function: The basic goal of our policy intends to seek a sequence of actions on the state space so as to find a reliable tracking routing. Moreover, we employ a designed consistency-check reward function to efficiently evaluate the reliability of the tracking results. Therefore, the objective function is formulated motivated by the deterministic policy gradient [22] as the following expectation form :

$$\begin{aligned} J(\pi) &= \int_{\mathcal{S}} \rho_\pi(\boldsymbol{s}) r\left(\boldsymbol{s}, f_\pi(\boldsymbol{s})\right) \mathrm{d}\boldsymbol{s} \\ &= \mathbb{E}_{\boldsymbol{s}\sim\rho_\pi}[r(\boldsymbol{s}, f_\pi(\boldsymbol{s}))], \end{aligned} \tag{1}$$

where $\mathbb{E}_{\boldsymbol{s}\sim\rho_\pi}[r(\boldsymbol{s}, f_\pi(\boldsymbol{s}))]$ denotes the expected value with respect to the discounted state distribution $\rho_\pi(\boldsymbol{s})$, and $f_\pi(\cdot)$ deterministically specifies our policy network.

2.2. Optimization

To optimize (1), we collect all weights of the CNN_{θ_π} for the parameters of both our policy network and critic network. Starting from a given state $\boldsymbol{s}_i$ and taking an action $\boldsymbol{a}_i$ under the policy π thereafter (i is the i-th iteration), we define the reliability critic function by the following Bellman equation:

$$Q_\pi(\boldsymbol{s}_i, \boldsymbol{a}_i) = \mathbb{E}[r(\boldsymbol{s}_i, \boldsymbol{a}_i) + \gamma \cdot Q_\pi(\boldsymbol{s}_{i+1}, f_\pi(\boldsymbol{s}_{i+1}))]. \tag{2}$$

where $\gamma \in [0, 1]$ is leveraged to smoothly weaken the intense dependency on previous iterations. Note that the expectation objective of learning policy takes advantages of integrating on only the state space during the training process [22].

Since our forward-cycle agent and backward-cycle agent are inferred based on inter-cycle communication via multi-agent policy gradient. Specifically, each approximate policy is learned by maximizing the log probability of the actions emitted by the other agent with an entropy regularizer [23]. Motivated by [22], the gradient of each policy is performed by minimizing the following optimization problem as:

$$J(\pi) = \mathbb{E}_{\boldsymbol{s},\boldsymbol{a},i}[(\mu_Q(\boldsymbol{s}_i, \boldsymbol{a}_i) - y_i)^2], \tag{3}$$

where the target value is computed based on (2) as

$$y_i = r(\boldsymbol{s}_i, \boldsymbol{a}_i) + \gamma \cdot \mu_Q(\boldsymbol{s}_{i+1}, \boldsymbol{a}_{i+1}). \tag{4}$$

To further enhance the training convergence, we employ a framework for prioritizing experience [24] for the efficient exploration, which typically replays significant transitions more frequently in the training phase.

3. EXPERIMENTS

To evaluate the effectiveness of our DCDRL, we conducted the main experimental results on the widely-used video-based face alignment dataset. Next, we present details on evaluation datasets, protocol and experimental analysis, respectively.

Table 1. Comparisons of averaged errors of our proposed DCDRL with the state-of-the-arts (68-lms, in chronological ranking). We found that our method with full-supervision signals gains best performance and even achieves compelling results with partial annotations with state-of-the-art approaches (in the chronological order).

Methods	Cate-1	Cate-2	Cate-3
SDM [4] (2013)	7.41	6.18	13.04
TSCN [26] (2014)	12.54	7.25	13.13
CFSS [5] (2015)	7.68	6.42	13.67
TCDCN [27] (2016)	7.66	6.77	14.98
CCR [28] (2016)	7.26	5.89	15.74
iCCR [28] (2016)	6.71	4.00	12.75
TSTN [13] (2018)	5.36	4.51	12.84
FHR [29] (2018)	4.12	4.18	5.98
DCDRL	**3.95**	**4.00**	**5.42**
Semi-DCDRL^{-50}	**4.33**	**4.31**	**5.68**
Semi-DCDRL^{-25}	**4.87**	**4.46**	**6.10**
Semi-DCDRL^{-10}	**4.96**	**4.57**	**6.36**

$\star$ $^{-50}$ $^{-25}$ $^{-10}$ denote the 50%, 25%, 10% of annotations were employed for training with DCDRL, respectively.

Evaluation Dataset and Protocol: The 300 Videos in the Wild (300-VW) Dataset [25] was collected typically for video-based face alignment, which contains 114 videos that were captured in various conditions and each video has around 25-30 images per second. By following the settings in [25], we utilized 50 sequences for training and the remaining 64 sequences were used for testing. Moreover, the whole testing set is divided into three categories (1, 2, 3): well-lit, mild unconstrained and challenging. Hence, the Category 3 directly exploits the difficult cases of face sequences, which highlights the superiority of the proposed approach. It is valuable to notified that we utilized 300-W [21] training set to initialize the policy network and critic network. For standard evaluation metric, we employed the normalized root mean squared error (RMSE) and cumulative error distribution (CED) curves in our experiments. We averaged the RMSEs of all frames within each category and then average them as final performance. Besides, we leveraged the CED curves [4,5] of RMSE errors to quantitatively evaluate the performance in comparisons to the state-of-the-arts.

Implementation Details: For the input data preparation, we detected faces on the whole dataset by enlarging the groundtruth annotations. Then we rescaled both the detected facial images with padding zeros and the corresponding annotations with the restricted output scales. For each evaluation dataset, the image resolution of the input is determined

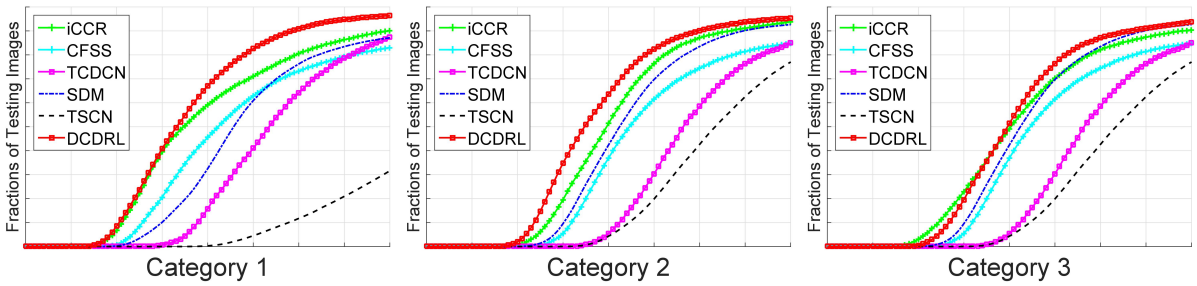

Fig. 3. CED curves of our DCDRL compared to the state-of-the-arts on all three categories in 300-VW [25], respectively, where standard 68 landmark were employed for evaluation. Our proposed DCDRL significantly outperforms state-of-the-art methods.

by the averaged size over all detected faces. In terms of the specification of the policy network, the first convolution layer is fed with L raw local patches in size of 26×26. The following two convolutional layers (3×3 kernel size, 1×1 stride) are in size with 64 and 128 kernels. Finally, we appended a two-layer fully connections parameterized by 128×256 and $256\times2L$ matrices. For hyper-parameters employed in our DCDRL, we empirically specified the discounted factor to 0.9 and the learning rate γ to 0.001, respectively. Besides, we sampled 100 transitions in the replay buffer.

Results and Analysis: We compared our approach with the state-of-the-art face alignment methods, which were designed for both still images and tracking in videos. For fair comparisons, we first leveraged all annotations for model training by following the common fully-supervised methods. To further highlight the advantage of our approach, we trained our model termed *Semi-DCDRL* by only a subset of the provided labels. Fig. 3 shows the CED curves of our method compared with the state-of-the-arts, where partial results are presented in Table 1. From these results, we see that our proposed DCDRL significantly outperforms other face alignment methods by a large margin, which is because our dual-cycle execution exploits more cues to learning discriminative spatial-temporal features for robust face tracking.

Seeing from Table 1 which tabulates the comparisons of DCDRL versus semi-DCDRL, we see that even with weakly-labeled annotations, our method degrades slightly even on the challenging cases due to large poses, diverse expressions and severe occlusions. This also demonstrates the effectiveness of the proposed dual-cycle modeling of the bidirectional orders, where these learned temporal cues are helpful to promote the stabilization for face tracking. Besides, to qualitatively visualize the results of our method compared with SBR [19], we selected a challenging 517-th video clip on the 300-VW dataset and performed results on all frames. As these curves shown in Fig. 4, we achieve a clear and stable results by a large margin compared with SBR [19], which shows the stabilization of ours versus various temporal motions.

Computational Time: During the testing phase, the forward agent performs tracking landmarks, while the backward agent learns to justify the tracking drifts for previous frames. Moreover, the detector will re-initialize our tracker when the RMSE error reaches higher beyond the threshold 0.01. In terms of efficiency performance, the whole training procedure requires 12 hours with a single NVIDIA TITAN V GPU card. Our model runs nearly at 23 frames per second on the Intel Xeon (R) Gold 5118 CPU@2.30Ghz platform.

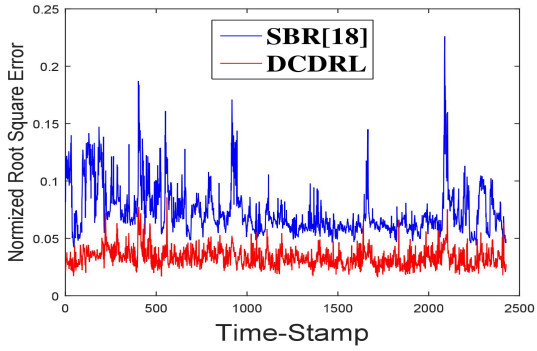

Fig. 4. Qualitative results of our proposed DCDRL compared with SBR [19] on the challenging 517-th sequence of the 300-VW Cate-3 dataset. We see that our model achieves low errors across all frames.

4. CONCLUSION

We have proposed a dual-cycle deep reinforcement learning method to address the stabilization for face tracking by using weak-supervision signals. Our architecture achieves two bidirectional temporal orders by accumulating plausible consistency-check rewards. Experimental results have manifested the effectiveness of our approach versus many difficult cases due to various temporal motions and occlusions. In future works, it is desirable to exploit multi-view faces in our approach to improve the robustness versus large poses in videos and moreover tackle the problem of personalized face tracking in a unified deep reinforcement learning framework.

5. REFERENCES

[1] Hu, J., Lu, J., Tan, Y.: Discriminative deep metric learning for face verification in the wild. In: CVPR. (2014) 1875–1882

[2] Grewe, C.M., Zachow, S.: Fully automated and highly accurate dense correspondence for facial surfaces. In: ECCVW. (2016) 552–568

[3] Cao, X., Wei, Y., Wen, F., Sun, J.: Face alignment by explicit shape regression. In: CVPR. (2012) 2887–2894

[4] Xiong, X., la Torre, F.D.: Supervised descent method and its applications to face alignment. In: CVPR. (2013) 532–539

[5] Zhu, S., Li, C., Loy, C.C., Tang, X.: Face alignment by coarse-to-fine shape searching. In: CVPR. (2015) 4998–5006

[6] Trigeorgis, G., Snape, P., Nicolaou, M.A., Antonakos, E., Zafeiriou, S.: Mnemonic descent method: A recurrent process applied for end-to-end face alignment. In: CVPR. (2016) 4177–4187

[7] Jourabloo, A., Ye, M., Liu, X., Ren, L.: Pose-invariant face alignment with a single cnn. In: ICCV. (Oct 2017)

[8] Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., Zhou, Q.: Look at boundary: A boundary-aware face alignment algorithm. In: CVPR. (2018)

[9] Liu, H., Lu, J., Guo, M., Wu, S., Zhou, J.: Learning reasoning-decision networks for robust face alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), accepted (2018)

[10] Wang, X., Yang, M., Zhu, S., Lin, Y.: Regionlets for generic object detection. TPAMI **37**(10) (2015) 2071–2084

[11] Guo, M., Lu, J., Zhou, J.: Dual-agent deep reinforcement learning for deformable face tracking. In: ECCV. (2018) 783–799

[12] Tzimiropoulos, G.: Project-out cascaded regression with an application to face alignment. In: CVPR. (2015)

[13] Liu, H., Lu, J., Feng, J., Zhou, J.: Two-stream transformer networks for video-based face alignment. TPAMI **40**(11) (2018) 2546–2554

[14] Peng, X., Feris, R.S., Wang, X., Metaxas, D.N.: A recurrent encoder-decoder network for sequential face alignment. In: ECCV. (2016) 38–56

[15] Doersch, C., Gupta, A., Efros, A.A.: Unsupervised visual representation learning by context prediction. In: ICCV. (2015) 1422–1430

[16] Wei, D., Lim, J.J., Zisserman, A., Freeman, W.T.: Learning and using the arrow of time. In: CVPR. (2018) 8052–8060

[17] Meister, S., Hur, J., Roth, S.: Unflow: Unsupervised learning of optical flow with a bidirectional census loss. In: AAAI. (2018) 7251–7259

[18] Jourabloo, A., Liu, X.: Large-pose face alignment via cnn-based dense 3d model fitting. In: CVPR. (2016) 4188–4196

[19] Dong, X., Yu, S., Weng, X., Wei, S., Yang, Y., Sheikh, Y.: Supervision-by-registration: An unsupervised approach to improve the precision of facial landmark detectors. In: CVPR. (2018) 360–368

[20] Chang, C., Chou, C., Chang, E.Y.: CLKN: cascaded lucas-kanade networks for image alignment. In: CVPR. (2017) 3777–3785

[21] Sagonas, C., Antonakos, E., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: 300 faces in-the-wild challenge: database and results. IVC **47** (2016) 3–18

[22] Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., Riedmiller, M.A.: Deterministic policy gradient algorithms. In: ICML. (2014) 387–395

[23] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., Mordatch, I.: Multi-agent actor-critic for mixed cooperative-competitive environments. In: NIPS. (2017) 6382–6393

[24] Schaul, T., Quan, J., Antonoglou, I., Silver, D.: Prioritized experience replay. In: ICLR. (2016)

[25] Shen, J., Zafeiriou, S., Chrysos, G.G., Kossaifi, J., Tzimiropoulos, G., Pantic, M.: The first facial landmark tracking in-the-wild challenge: Benchmark and results. In: ICCVW. (2015) 1003–1011

[26] Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. In: NIPS. (2014) 568–576

[27] Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Learning deep representation for face alignment with auxiliary attributes. TPAMI **38**(5) (2016) 918–930

[28] Sánchez-Lozano, E., Martínez, B., Tzimiropoulos, G., Valstar, M.F.: Cascaded continuous regression for real-time incremental face tracking. In: ECCV. (2016) 645–661

[29] Tai, Y., Liang, Y., Liu, X., Duan, L., Li, J., Wang, C., Huang, F., Chen, Y.: Towards highly accurate and stable face alignment for high-resolution videos. AAAI, in press (2019)