Reinforced Learning for Label-Efficient 3D Face Reconstruction

Hoda Mohaghegh[1], Hossein Rahmani[2], Hamid Laga[3], Farid Boussaid[4], Mohammed Bennamoun[1]

***Abstract*— 3D face reconstruction plays a major role in many human-robot interaction systems, from automatic face authentication to human-computer interface-based entertainment. To improve robustness against occlusions and noise, 3D face reconstruction networks are often trained on a set of in-the-wild face images preferably captured along different viewpoints of the subject. However, collecting the required large amounts of 3D annotated face data is expensive and time-consuming. To address the high annotation cost and due to the importance of training on a useful set, we propose an Active Learning (AL) framework that actively selects the most informative and representative samples to be labeled. To the best of our knowledge, this paper is the first work on tackling active learning for 3D face reconstruction to enable a label-efficient training strategy. In particular, we propose a Reinforcement Active Learning approach in conjunction with a clustering-based pooling strategy to select informative view-points of the subjects. Experimental results on 300W-LP and AFLW2000 datasets demonstrate that our proposed method is able to 1) efficiently select the most influencing view-points for labeling and outperforms several baseline AL techniques and 2) further improve the performance of a 3D Face Reconstruction network trained on the full dataset.**

I. INTRODUCTION

Monocular 3D face reconstruction enables a wide range of computer vision applications in face recognition, human–computer interactions, virtual/augmented reality and autonomous driving [1]. Recently, deep learning based 3D face reconstruction methods [2], [3], [4], [5], [6], [7], [8], [9] have demonstrated significant success due to improved representation power. However, such models need to be trained on large-scale 3D training datasets. Such datasets are extremely time-consuming and expensive to collect and annotate. To achieve effective 3D face reconstruction given a limited annotation budget, we resort to selecting a subset of informative examples as training data and sending them to a human oracle to be annotated. The process will be repeated until the termination criterion (e.g., when the annotation budget is exhausted) is met. Specifically, we propose a reinforced active learning framework in which by selecting the most influencing and representative samples to label, the reconstruction algorithm can achieve a high performance with a minimum number of annotated face images. We investigate the scenario of: *"Which viewpoint per subject is more informative?"* and based on that, we define the corresponding pooling and selection strategy as follows. As each query subject arrives, we perceive its corresponding viewpoints as the unlabeled gallery pool. Here, we aim to discard the misleading and confounding viewpoints and find the most informative ones among multiple possible views per subject to train the 3D face reconstruction network. In most existing works, the core of the AL-based methods, i.e., sampling unit, uses some heuristic selection methods to maximize the informativeness and representativeness of the selected samples [10], [11], [12], [13], [14], [15], [16]. Rather than manually defined heuristics, we leverage a Reinforcement Learning model in which an active learner *learns* the sampling policy in a data-driven manner. In the proposed framework, the view-point selection decision is made based on 3D reconstruction model's prediction and its uncertainty to account for the informativeness of the selected samples. In addition, we enforce diversity, to avoid redundancy, using a clustering-based pooling in our AL framework. Clustering unlabeled data based on the face identity features ensures broad coverage over the entire data. By minimizing the error metric between the estimated and the ground truth 3D face, the RL agent is trained to find the most informative and discriminative samples over a set of unlabeled face images.

In summary, our contributions are: **(1)** We propose a label-efficient learning strategy for 3D face reconstruction under Active Learning framework. **(2)** In our proposed framework, for the first time we successfully learn a Reinforcement Learning-based acquisition function as a sampling strategy such that a 3D face reconstruction network can achieve a high performance with a minimum number of labeled data. **(3)** The proposed pooling strategy in conjunction with the model uncertainty leveraged in our RL agent training process allows the sampling strategy to exploit both representativeness and informativeness. **(4)** With the extensive experiments on both 3D face reconstruction and face alignment (landmark detection) tasks on AFLW2000 and 300W-LP datasets [9], we demonstrate the superiority of the proposed learning strategy over competitive AL baselines with significant performance gain whilst using much less annotations. Particularly, on 300W-LP and AFLW2000, our method achieves nearly the same performance as the 3D reconstruction model trained on the whole training set, using less than 30% and 40% of data, respectively.

II. RELATED WORK

A. 3D Face Reconstruction

Various approaches have been proposed to tackle the inherently ill-posed problem of 3D face reconstruction from a single image; see [17] for a detailed survey. The biggest obstacle to applying deep networks to 3D face reconstruction

[1]Hoda Mohaghegh and Mohammed Bennamoun are with the school of Computer Science and Software Engineering, University of Western Australia.

[2]Hossein Rahmani is with the Department of Computing and Communications, Lancaster University.

[3]Hamid Laga is with the Information Technology Discipline, Murdoch University.

[4]Farid Boussaid is with the School of Engineering, Electrical, Electronic and Computer Engineering, University of Western Australia.

lies in the lack of training data. This is because collecting a large amount of 2D face images together with the corresponding 3D ground truth required by deep learning-based approaches is both time and cost consuming.

For supervised methods, the ground-truth 3D geometry of human faces can be generated by time-consuming optimization-based methods, such as Gaussian Process [18] and Parameterized Spline [19]. For face images in the wild, which would exhibit occlusions, non-uniform lighting and cluttered backgrounds, such methods cannot guarantee accurate ground-truth geometry without human intervention. Consequently, we cannot easily collect a large amount of 3D annotated training data, and the reconstruction accuracy of supervised methods is thus restricted by the inadequate amount of training data.

Another solution is to focus on un/self-supervised learning. Recently, several works proposed to supervise the reconstruction procedure by utilizing the reconstruction (photometric) or adversarial loss on the rendered images without using any explicit 3D annotations [20], [21], [22]. However, in fully unsupervised approaches, calculating such losses requires simultaneously estimating both shapes and textures. Given that the initial estimation is far from ideal to render meaningful images, the unfaithful reconstructions might be obtained. Also, such methods suffer from the depth-scale ambiguity and thus may predict an incorrectly scaled face [2].

B. Active Learning in Computer Vision Tasks

Active Learning is a well-studied research domain applied to several tasks, e.g., semantic segmentation [23], [24], [25], image classification [13], [26], [27], 3D hand pose estimation [28], [29], human pose estimation [30], [31] and natural language processing [32], to reduce the data labeling effort. Ren et al. [33] provides a recent survey of various standard active learning methods for various tasks. The goal of AL is to obtain satisfactory performance for the model at the smallest possible labeling cost. In other words, given a machine learning model and a pool of unlabeled data, instead of asking human to label all the unlabeled data, AL iteratively selects which samples should be labeled next. Existing AL approaches have examined both pooling-based [28], [29], [31], [30] and stream-based [32], [34] strategies. In the former class, datapoints are selected from a large pool of unlabeled data and in the later one, the unlabeled samples are provided one by one and the decision is to label it or not.

In this paper, we propose a pooling-based active learning framework for 3D face reconstruction to enable a label-efficient learning strategy. To the best of our knowledge, there has been no study of the applicability of AL to 3D reconstruction, especially for 3D facial data. Rather than using manually-designed heuristics, we propose a reinforcement active learner that learns the sampling policy from data. Unlike most of existing uncertainty-based sampling techniques, which directly query for most ambiguous samples, we leverage uncertainty as a measure of informativeness, in a Markov Decision Process. Our framework differs from other existing methods by the task we tackle, the formulation of our sampling space, and details of the reinforcement active learner we employ to find the optimal policy.

III. Methodology

In this section, we first outline the overall pipeline of our proposed framework for label-efficient 3D face reconstruction and then we elaborate on different components of the pipeline. Our goal is to train a 3D face reconstruction network on a minimum number of view-points per person while maximizing its performance. To this end, a query network (an agent) selects the most informative samples among a pool of unlabeled data, which are then annotated by a human oracle. These samples are added to the labeled set used to train the supervised 3D reconstruction network. This process is done iteratively until a given annotation budget is achieved.

The active learner in our framework is formalized as a Markov Decision Process (MDP), which allows the learning of a Reinforcement Learning agent. We adopt the Q-learning algorithm [35] to solve this MDP problem resulting in intelligent selection of the most informative samples by a policy network. In what follows, we detail the mechanism for learning the policy network yielding a labeling decision.

A. Overall Framework

As can be seen in Fig 1, given a set of unlabeled samples, our method first clusters the unlabeled face images based on their identities with each cluster containing a single subject along with its various view-points. Thus, as each query subject arrives, all of its corresponding view-points together with the query data are considered as our unlabeled gallery pool. Instead of selecting one subject per iteration, which is highly inefficient since each iteration involves updating the 3D face reconstruction network and computing the reward, we propose to select a number of subjects K_{sub} in each iteration. At each time step t, the environment provides an observation (state-action representation), which describes the relationship between samples, using the concatenation of the current network's prediction and its uncertainty. It receives a response from the agent by providing scores for candidates in the pool and selecting an action. The agent requests the sample with the highest score among the unlabeled gallery pool being annotated by the human oracle. Instead of a human oracle, as common in active learning approaches, we mask out the ground truth of the fully labeled dataset and reveal them when the active learning algorithm selects them to be annotated. However, in real-life applications of our AL framework, we can apply it in a setting with unlabeled data with a human in the loop labeling the selected data.

When sufficient viewpoints from a number of subjects are obtained, the labeled dataset will be updated and the 3D face reconstruction model's (here, PRN [6]) parameters are updated, which in return generates a new trained network for computing state and action representations. Then, the agent receives a reward based on the improvement in the performance of the 3D face reconstruction network trained

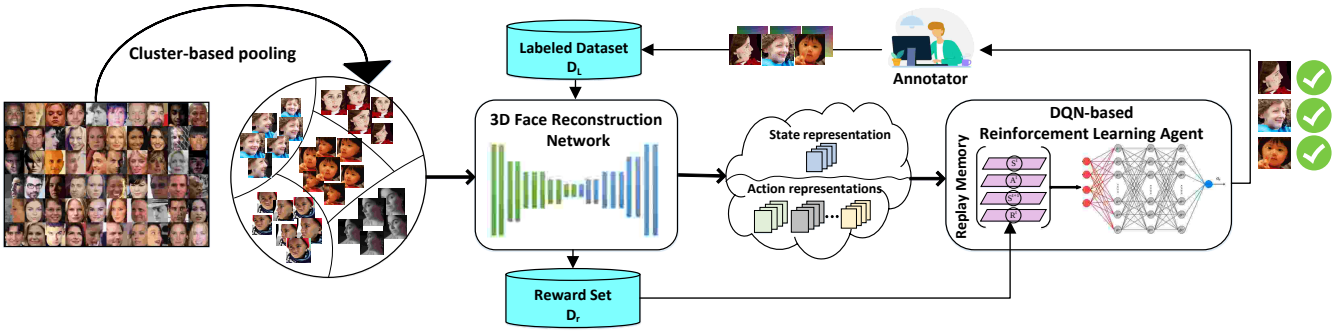


Fig. 1. Proposed deep reinforcement learning in which an agent is designed to dynamically select the most informative instances for 3D Face Reconstruction network.

with the selected samples. All subjects will be browsed once and the whole process terminates when the pre-defined budget is reached. In summary, during training, the algorithm iterates over each subject to play the best viewpoint selection game. Each iteration of the game consists of the following steps:

1. Computing the state-action representation (s_t, a_t) on query and candidates as a function of PRN (Section 3.2).

2. Selecting actions A_t by following the scores that a policy network Π generates. The actions define the K_{view} viewpoints per subject with the highest score to be annotated.

3. Updating the labeled and unlabeled set by adding the selected viewpoint and its ground truth to the labeled set i.e., $D_L^{t+1} = D_L^t \cup (x_j, y_j), j = 1, \ldots, K_{sub} \times K_{view}$ and removing them from unlabeled set i.e., $D_U^{t+1} = D_U^t - (x_j, y_j), j = 1, \ldots, K_{sub} \times K_{view}$.

4. Training the 3D face reconstruction network using the updated labeled set with the recently added viewpoints D_L^{t+1} to generate PRN^{t+1}.

5. Computing the reward R^t as the difference of 3D face reconstruction's performance between PRN^{t+1} and PRN^t on the reward set (D_r).

More details regarding the notations and the entire algorithm can be found in Algorithm 1. Below, we elaborate on the detailed definition of the state-action representation, reward and policy learning.

B. State-Action Representation

In order to find the optimal policy, the agent interacts with the environment and receives data, which is used during the training of the Q-Network. Q-Networks in general, map environment states to agent actions that maximize the expected sum of rewards [36]. To help the agent to make the best decision and select the most informative samples, the proposed state should ideally characterize the distribution gap between the labeled data and the unlabeled data at iteration t. On the other hand, action should represent the contribution of the candidate unlabeled sample when it is selected to be labeled. Here, we rely on a Deep Q-Network, which takes the state-action representation as input and returns a single value as the score of each action. In other words, each action A_t is associated with a candidate view-point in our clustered unlabeled pool. To assist the decision process, we applied the combined state-action representation at time t as it describes the relationship between the candidate viewpoints being considered for annotation, the query image and the labeled dataset constructed up to time t. This relationship is represented by the ensemble of two different feature sets: one describes the network's prediction and the other one approximates the model uncertainty.

The first set of features is directly extracted from the output of our 3D face reconstruction network, i.e., PRN [6], trained on the current labeled dataset. The second set of features, i.e., epistemic uncertainty, accounts for the uncertainty in the model parameters. Many AL acquisition functions directly utilize model uncertainty to select the most ambiguous samples. Unlike existing uncertainty-based approaches, we deployed this informative cue as an observation of the environment for the policy network to make its decision. However, modeling the uncertainty in regression problems is not as straightforward as for classification tasks whose uncertainty is estimated by the posterior probability of a class [37]. To model the uncertainty in our regression problem, we first make our 3D face reconstruction network a probabilistic model by an approximation of a Bayesian Neural Network. Particularly, we use the Monte Carlo Dropout (MCD) technique [38] to obtain an approximation of the posterior's mean and variance.

In practice, by applying MCD, we obtained a Bayesian model in which by L times passes of a sample, our first set of features is calculated as follows:

$$\mu_{pred} = \frac{1}{L} \sum_{l=1}^{L} y_l, \tag{1}$$

where y_l is the model's prediction at l^{th} time and μ_{pred} is the average of predictions over L forward passes through the network. In the Bayesian approximation, the second set of features, i.e., epistemic uncertainty, can be evaluated using variational inferences as follows:

$$\sigma_{epis} = \frac{1}{L} \sum_{l=1}^{L} y_l^2 - \mu_{pred}^2. \tag{2}$$

To summarise, the final state-action representation at time t, is the concatenation of these two sets of features (μ_{pred} and σ_{epis}) for the query sample and candidate view-points. To avoid intensive memory usage, we need to downsample our state-action representation pair. Finally, the agent scores each state-action representation pair (S^t, A^t) corresponding to the unlabeled viewpoint x_i and takes the action A^t with the

Algorithm1. Learning an active learning policy

Input: D_U, D_L, D_r, N, N_v, K_{sub}, K_{view}, B, $Pre-trained\ PRN$
Output: Trained policy network π

for episode $i = 1, 2, \dots, N$ **do**:
  Reload $Pre-trained\ PRN$, $D_L \leftarrow \emptyset$ and shuffle clustered D_U
  while the labeling budget (B) is not spent, **do**:
    Select K_{sub} query samples and their N_v corresponding viewpoints to build the pool.
    For each subject, score each viewpoint using the policy net and the computed state-action representations S^t, A^t.
    Sort the viewpoints based on the scores and select K_{view} viewpoints with the highest scores for each subject.
    Update D_L: $D_L = D_L + K_{view}\ selected\ views$
    Update D_U: $D_U = D_U - K_{view}\ selected\ views$
    Train PRN on updated $D_L \rightarrow PRN^{t+1}$
    Compute the reward on $D_r \rightarrow R^t$
    Select K_{sub} new subjects and their corresponding viewpoints to build the new pool.
    Compute state representation S^{t+1}
    Add the (S^t, A^t, S^{t+1}, R^t) to the experience buffer.
    Use the standard DQN algorithm to optimize the policy network using the experience dataset.
  end while
end for
return trained policy network π

highest score. When the informative viewpoint is chosen, an oracle is requested to label the sample. The newly annotated sample is added to the training data and PRN is subsequently updated.

C. Reward

Each time the 3D face reconstruction network is trained on the recently added viewpoints, the agent receives a reward, which provides feedback on the quality of the actions made by the agent. Here, the reward is defined as the change in the performance of the 3D face reconstruction model, i.e., $R(s_t, a) = Error(PRN^{t+1}) - Error(PRN^t)$ on D_r set. In this equation, $Error$ denotes the normalized Mean Error between the outputs of 3D Face Reconstruction network (here, UV-position maps) and PRN^{t+1} is the trained model after action a has taken place.

D. RL-based Policy Learning

In our Active Learning framework, we adopt a Reinforcement Learning approach to learn an optimal policy using the above-mentioned components. In particular, we formulate the problem of finding an optimal policy, which maps a state into an appropriate action (i.e., choosing the appropriate viewpoints) as a Markov Decision Process (MDP). This process is described by (s_t, a_t, r_t, s_{t+1}) which stand for the states, actions, rewards and next states, that the agent turns to through the actions. Following Mnih [39], we adopt a technique known as experience replay where we store the agent's experiences at time t, (s_t, a_t, r_t, s_{t+1}) into a replay memory M. A mini-batch of transitions from an experience buffer is then sampled randomly and will be fed to the Deep Q-Network to generate $Q^\pi(s, a)$. Training DQN is basically a regression problem where the objective is to match Q-values predicted by DQN and the expected (target) Q-values from the Bellman equation: $r_i + \gamma max_a Q(s_{i+1}, a)$, where r is the immediate reward and γ is the discount factor which controls the contribution of rewards. Following Mnih [39], we performed the stochastic gradient descent on the loss function:

$$L(\theta) = (y_j - Q(s_j, a_j))^2. \quad (3)$$

Here, $y_j = r_j + \gamma max_{a\prime} Q(s_{j+1}, a\prime)$ is the target Q-value based on the current parameters. There are two approaches to design the architecture of the deep Q-network. In the standard one proposed in [39], the Q-function takes the state representation as the input and generates the Q-value of all possible actions. The second approach takes the state and the action as the input to the Q-function and adapts the standard DQN architecture accordingly to produce the single Q-value of the action. We adopt the second architecture to design our DQN due to the different numbers of viewpoints per subject (i.e., different number of possible actions for each subject or query image).

IV. Experiments

We conducted extensive experiments on 3D face reconstruction and dense alignment to evaluate the performance of the proposed label-efficient framework.

A. Experimental Setting

We trained our active learning framework on 300W-LP [9], as it is the only publicly available dataset containing face images across different angles with fitted 3DMM parameters. As explained in Section 3, we split 300W-LP into the five subsets D_{init}, D_{rew}, D_{val}, D_{tr} and D_{test}, which contain approximately 25K, 2K, 3K, 30K and 1225 images, respectively. In addition to D_{test}, we used AFLW2000 dataset [9] to evaluate the performance of the proposed AL-based framework. In our AL-based framework, we use the Position map Regression Network (PRN) [6], a very light weight model, to jointly predict dense alignment and reconstruct 3D face shape. As our proposed model consists of a deep active learning framework on the top of a 3D face reconstruction module, it can be easily adjusted for other 3D face reconstruction networks.

For quantitative evaluation, we used NME2D and NME3D, which represent the Normalized Mean Error (NME) between the ground truth UV-position map and predicted one in 2D and 3D spaces, respectively. We also evaluated the face alignment performance using NME on a sparse set of 68 facial landmarks in which the bounding box size is used as the normalization factor. The annotation budget size, B, the number of subjects to select, K_{sub}, and the number of viewpoints per subject to select, K_{view}, in each AL iteration were set to 27000, 50 and 1, respectively.

B. Competing Methods

We compared the proposed method against the following seven baselines in AL approaches. The first four are model-free sampling strategies and the next three are model-based algorithms that have been used in most of the comparative analyses in the literature. Since there is no prior work on active learning for 3D Face Reconstruction, we explored the following AL strategies proposed in other active learning domains and adapted them to our task. The details of these approaches are described as follows:

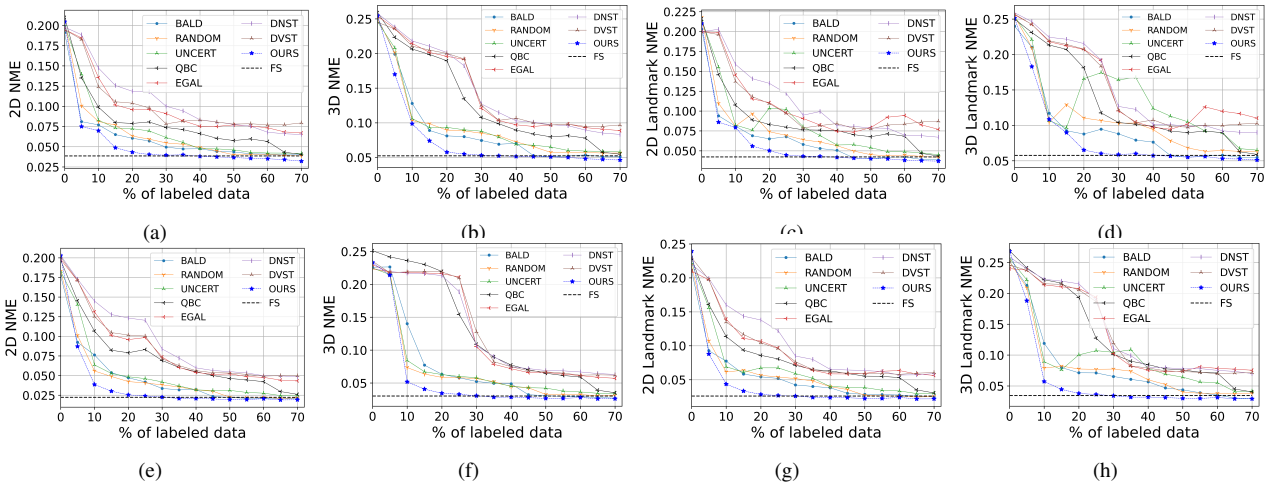


Fig. 2. Quantitative comparison of different AL baselines on AFLW2000 dataset (a)-(d) and 300W-LP dataset (e)-(h). The results of 3D Face Reconstruction are presented in (a), (b), (e), (f) and the results of landmark detection are presented in (c), (d), (g) and (h). Lower is better in all sub-figures.

Random Sampling (RND) is a typical approach for annotating data by uniformly sampling the unlabeled pool.

Exploration Guided Active Learning (EGAL) is an example of a density-weighted approach, which also takes into account the representativeness of each instance of the dataset as a whole. The choice of the balancing parameter ω was shown to play an important role in the effectiveness of the EGAL selection strategy. For $\omega = 0.25$, the EGAL algorithm was shown to consistently outperform other sampling strategies [14].

Pure Diversity (DVST) with $\omega = 0$, EGAL results in a purely diversity-based sampling as only the most diverse example is added to the candidate set; and so will be selected regardless of density [14].

Pure Density (DNST) with $\omega = 1$, EGAL results in a density-only approach as all unlabelled examples are added to the candidate set, regardless of their diversity score [14].

Uncertainty Sampling (UNCERT) is a widely used selection criteria, which selects the topmost uncertain data with maximum cumulative epistemic variances [40].

Bayesian Active Learning by Disagreement (BALD) is a sampling technique, which selects a data point that is expected to maximise the information gained about the model parameters, i.e., maximise the mutual information between predictions and model posterior [13].

Query By Committee (QBC) is an active learning approach, which builds a committee of learners from existing labeled training data set and queries the instances that cause maximum disagreement among the committee. We have tested the QBC algorithm with different numbers of committees e.g., 3, 4, 5 and 7 and reported its best performance i.e., using 7 committees [15].

C. Comparison Results

We quantitatively evaluated our proposed framework for label-efficient 3D face reconstruction and face alignment (facial landmark detection). To have a fair comparison with other AL methods, the same pooling strategy was adopted before applying active learning techniques. Fig. 2 shows the various quantitative results in terms of the four different evaluation metrics described in Section IV-A. Figs. 2 (a-d) show the error values as a function of the percentage of annotated data on AFLW 2000 dataset. We continue the curves until 70% of all initially unlabeled data has been labelled. The PRN achieves 0.038, 0.052, 0.042, and 0.057 for 2D NME, 3D NME, 2D landmark NME and 3D landmark NME, respectively, by being trained with all labeled data in 300W-LP. We obtain these errors using only less than 40% of labeled data. This shows the effectiveness of our proposed framework in reducing the labelling cost, which is important for the 3D reconstruction task. Fig. 2 also compares our proposed method against various AL baselines at different AL iterations. It can be observed that our approach outperforms the baselines by a clear margin for every fixed budget, except for 50% where we achieve a similar performance as BALD in Fig. 2 (b-d).

Fig. 2 (e-h) show quantitative results on 300W-LP test set for different budget sizes. Here, we also observe that our method outperforms the baselines for all budget points, which demonstrates the effectiveness of our sampling strategy. By labeling approximately 9K images, corresponding to only 30% of the total images, we obtained the performance of fully supervised PRN if it had access to all annotated data. To reach the same output, the closest method to us, i.e., BALD, requires more than 15K labeled images. The superiority of our method is more visible in the first iterations, for which we can achieve nearly the same result as a fully supervised model using only 20% of data. Other approaches require significantly more annotated data to achieve the same performance. From Fig. 2, it can be inferred that DVST and DNST, despite being less computationally intensive, have almost always the lowest performance on both datasets as they just consider the diversity and density of images in the sampling process. EGAL could improve their efficiency by combining diversity and density in an effective way. However, this model-free sampling approach still does not deliver satisfactory performance especially in the first iterations, while it can gradually achieve comparable one to other baselines when labeling larger fractions of the data.

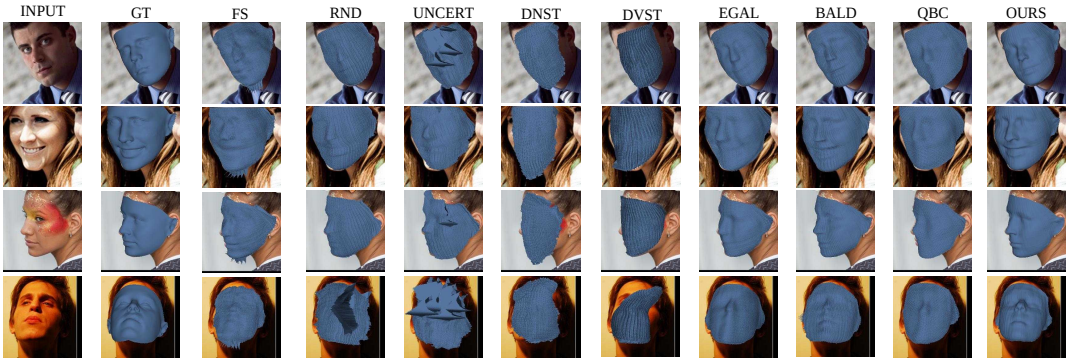


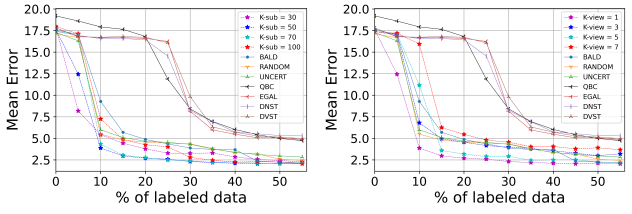
Fig. 3. 3D Face Reconstruction under different active selection strategies where the annotation budget is set as 50% (15K). The ground truth 3D meshes (GT) and reconstructed 3D meshes by fully supervised PRN, trained on 100% of data (FS), are shown in second and third columns, respectively. Using only 50 % of training data, selected by our proposed sampling method, the PRN is capable of recovering accurate 3D faces.

Rather than adopting a fixed heuristic selection strategy, our proposed method *learns* what kind of data points are most informative and beneficial for training the model using the current state of the trained network. In such a data-driven scenario, our method is able to efficiently select the most informative samples to label so that the 3D face reconstruction model can reach the best possible performance in a cost-effective way.

We also qualitatively evaluated our proposed AL framework on 3D face reconstruction. We present a comprehensive comparison with common active learning baselines in Fig. 3 where the annotation budget is set as 15K (50% of the whole dataset). In this figure, the ground truth and the results of the fully supervised PRN trained on 100% of data, are shown in column 2 and 3, respectively. It can be seen that our method outperforms all alternative active learning methods and the reconstructed meshes are closer to that of the ground truth. These characteristics are more visible for faces with extreme poses as shown for the third and forth input images. In particular, for faces with extreme poses and occlusions, the PRN trained on selected data by other AL methods has more difficulties to recover the detailed shapes as there are not sufficiently effective training samples for such poses. However, by selecting the informative viewpoints of each subject, our policy network is able to reconstruct faithful 3D shapes.

D. Ablation Study

Effect of the AL parameters. To analyse the effect of user-defined parameters (e.g., the number of subjects per iteration K_{sub} and the number of selected viewpoints per subject K_{view}) on our proposed framework, we report the results of our method using various K_{sub} and K_{view} values. In Fig. 4-a, we investigate the performance of our method under different K_{sub} on the 300W-LP dataset. Regardless of the choice of K_{sub}, our method is seen to consistently outperform almost all the other AL baselines with the best performance obtained when K_{sub} is set to 50.



(a) (b)

Fig. 4. Results of ablation tests on 300W-LP dataset using various (a) K_{sub} and (b) K_{view}. Smaller budget size allows our method to react faster to the training outcomes.

We also looked into the sensitivity of our proposed method to the number of selected viewpoints per subject by training the model under various K_{view}. From 4-b, it can be observed that by increasing the number of selected viewpoints from 1 to 7, the performance declines while the training time is reduced. However, they still outperform other AL baselines, which demonstrates the stability of our method for a wide range of budget sizes.

Effect of the state-action representations. We further performed an ablation study to evaluate the contribution of each component in our state-action representation. We investigated the influence of state-action components by individually excluding them from the full model. In Table I, we report the model performance for different active learning iterations on both AFLW2000 and 300W-LP datasets. It can be clearly concluded that all terms contribute to performance improvement and that our full model gives the lowest average NME error.

TABLE I

ABLATION STUDY ON THE VALIDITY OF TWO COMPONENTS IN OUR STATE-ACTION REPRESENTATION.

σ_{epis}	μ_{pred}	AFLW2000				300W-LP			
		3K	6K	9K	15K	3K	6K	9K	15K
✓	–	0.072	0.047	0.045	0.041	0.063	0.037	0.024	0.021
–	✓	0.076	0.051	0.042	0.042	0.060	0.034	0.032	0.024
✓	✓	**0.069**	**0.043**	**0.039**	**0.036**	**0.038**	**0.025**	**0.022**	**0.019**

V. CONCLUSION

In this paper, we took the first steps towards active learning for label-efficient 3D face reconstruction. We successfully employed a DQN-based reinforcement learning agent in the sampling unit of our proposed AL framework to select informative view-points and discard redundant and misleading ones. Under the proposed pool-based scenario, we achieve the lowest NME with the least amount of data for both 3D face reconstruction and facial landmark detection tasks on two well-known face datasets. We have shown that our proposed learning strategy outperforms competitive AL baselines and even the 3D face reconstruction model trained on the whole training set.

REFERENCES

[1] R. S. Siqueira, G. R. Alexandre, J. M. Soares, and G. A. Thé, "Triaxial slicing for 3-d face recognition from adapted rotational invariants spatial moments and minimal keypoints dependence," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3513–3520, 2018.

[2] W. Zielonka, T. Bolkart, and J. Thies, "Towards metrical reconstruction of human faces," *arXiv preprint arXiv:2204.06607*, 2022.

[3] E. Richardson, M. Sela, R. Or-El, and R. Kimmel, "Learning detailed face reconstruction from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1259–1268.

[4] Z. Bai, Z. Cui, X. Liu, and P. Tan, "Riggable 3d face reconstruction via in-network optimization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6216–6225.

[5] L. Tran, F. Liu, and X. Liu, "Towards high-fidelity nonlinear 3d face morphable model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1126–1135.

[6] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, "Joint 3d face reconstruction and dense alignment with position map regression network," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 534–551.

[7] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, "Large pose 3d face reconstruction from a single image via direct volumetric cnn regression," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1031–1039.

[8] A. Tuan Tran, T. Hassner, I. Masi, and G. Medioni, "Regressing robust and discriminative 3d morphable models with a very deep neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5163–5172.

[9] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3d solution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 146–155.

[10] Y. Yang, Z. Ma, F. Nie, X. Chang, and A. G. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 113–127, 2015.

[11] A. Vezhnevets, V. Ferrari, and J. M. Buhmann, "Weakly supervised structured output learning for semantic segmentation," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 845–852.

[12] A. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *2009 ieee conference on computer vision and pattern recognition*. IEEE, 2009, pp. 2372–2379.

[13] Y. Gal, R. Islam, and Z. Ghahramani, "Deep bayesian active learning with image data," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1183–1192.

[14] R. Hu, S. Jane Delany, and B. Mac Namee, "Egal: Exploration guided active learning for tcbr," in *International Conference on Case-Based Reasoning*. Springer, 2010, pp. 156–170.

[15] R. Gilad-Bachrach, A. Navot, and N. Tishby, "Query by committee made real," *Advances in neural information processing systems*, vol. 18, 2005.

[16] J. E. Iglesias, E. Konukoglu, A. Montillo, Z. Tu, and A. Criminisi, "Combining generative and discriminative models for semantic segmentation of ct scans via active learning," in *Biennial International Conference on Information Processing in Medical Imaging*. Springer, 2011, pp. 25–36.

[17] A. Morales, G. Piella, and F. M. Sukno, "Survey on 3d face reconstruction from uncalibrated images," *Computer Science Review*, vol. 40, p. 100400, 2021.

[18] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187–194.

[19] H. Guo, J. Jiang, and L. Zhang, "Building a 3d morphable face model by using thin plate splines for face reconstruction," in *Chinese Conference on Biometric Recognition*. Springer, 2004, pp. 258–267.

[20] B. Gecer, S. Ploumpis, I. Kotsia, and S. Zafeiriou, "Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1155–1164.

[21] A. Tewari, M. Zollhofer, H. Kim, P. Garrido, F. Bernard, P. Perez, and C. Theobalt, "Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 1274–1283.

[22] K. Genova, F. Cole, A. Maschinot, A. Sarna, D. Vlasic, and W. T. Freeman, "Unsupervised training for 3d morphable model regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8377–8386.

[23] Y. Siddiqui, J. Valentin, and M. Nießner, "Viewal: Active learning with viewpoint entropy for semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9433–9443.

[24] S. Sinha, S. Ebrahimi, and T. Darrell, "Variational adversarial active learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5972–5981.

[25] R. Sheikh, A. Milioto, P. Lottes, C. Stachniss, M. Bennewitz, and T. Schultz, "Gradient and log-based active learning for semantic segmentation of crop and weed for agricultural robots," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1350–1356.

[26] W. H. Beluch, T. Genewein, A. Nürnberger, and J. M. Köhler, "The power of ensembles for active learning in image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9368–9377.

[27] B. J. Meyer and T. Drummond, "The importance of metric learning for robotic vision: Open set recognition and active learning," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2924–2931.

[28] J. Gong, Z. Fan, Q. Ke, H. Rahmani, and J. Liu, "Meta agent teaming active learning for pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11 079–11 089.

[29] R. Caramalau, B. Bhattarai, and T.-K. Kim, "Active learning for bayesian 3d hand pose estimation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 3419–3428.

[30] B. Liu and V. Ferrari, "Active learning for human pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4363–4372.

[31] D. Yoo and I. S. Kweon, "Learning loss for active learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 93–102.

[32] M. Fang, Y. Li, and T. Cohn, "Learning how to active learn: A deep reinforcement learning approach," *arXiv preprint arXiv:1708.02383*, 2017.

[33] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, "A survey of deep active learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 9, pp. 1–40, 2021.

[34] M. Woodward and C. Finn, "Active one-shot learning," *arXiv preprint arXiv:1702.06559*, 2017.

[35] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[36] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3, pp. 279–292, 1992.

[37] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[39] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[40] D. Wang and Y. Shang, "A new active labeling method for deep learning," in *2014 International joint conference on neural networks (IJCNN)*. IEEE, 2014, pp. 112–119.