2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

DIAT (Depth-Infrared Image Annotation Transfer) for Training a Depth-Based Pig-Pose Detector

Steven Yik[1], Madonna Benjamin[2], Michael Lavagnino[3] and Daniel Morris[4]

***Abstract*— Precision livestock farming uses artificial intelligence to individually monitor livestock activity and health. Tracking individuals over time can reveal health indicators that correlate with productivity and longevity. For instance, locomotion patterns observed in lame pigs have been shown to correlate with poor animal welfare and productivity. Kinematic analysis of pigs using pose estimates provides a means of assessing locomotion. New dense depth sensors have potential to achieve full 3D pose estimation and tracking. However, the lack of annotated dense depth datasets has limited use of these sensors in detecting animal pose. Current annotation methods rely on human labeling, but identifying hip and shoulder locations is difficult for pigs with few prominent features, and is especially difficult in depth images as these lack albedo texture. This work proposes a solution to quickly generate high accuracy pig landmark annotations for depth-based pose estimation. We propose Depth-Infrared Annotation Transfer (DIAT), an approach that semi-automatically finds, identifies, and tracks marks visible in infrared, and transfers these labels to depth images. As a result, we are able to train a precise pig pose detector that operates on depth images.**

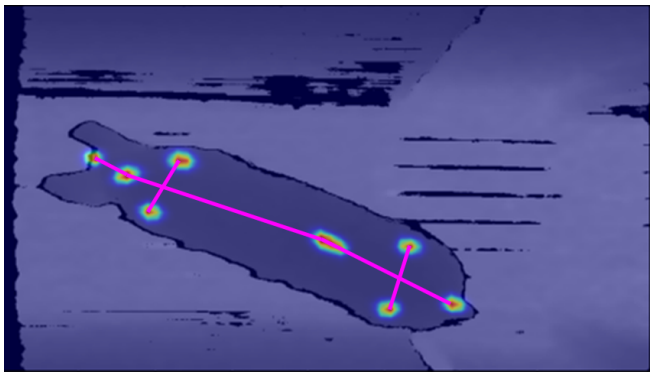

Fig. 1. Automatic detection of pig pose in a depth image by our trained network. Joint location heatmaps are shown as well as the resulting skeletal connections. In order to train this depth-based detector, we propose DIAT which uses landmarks visible in an IR image and transfers these to the depth image for network training.

I. INTRODUCTION

Precision livestock farming (PLF) aims to improve both animal welfare and farmer productivity as well as ease the burden on caregivers [34]. A critical technology enabling this is automated monitoring of individual animals. While advances have enabled coarse pose estimates of pigs [25], here our goal is to take the next step towards precise 3D tracking of pigs as they move around a farm. This will open the door to automated collection and analysis of shape and motion-based health metrics for pigs.

Research has demonstrated the potential of PLF metrics, such as body condition (thin, average, fat), and lameness (abnormal locomotion), as predictors of longevity and productivity [2, 13, 29, 32]. Currently, methods to measure body condition includes a human utilizing a caliper tool [15] or human observation of locomotion [6]. These modes of evaluation are prone to inconsistencies due to human error, transcription and subjectivity.

Advances in tracking human posture have been achieved through new sensing modalities such as the Kinect paired with machine learning [28], as well as the use of convolutional neural networks (CNNs) for human posture estimation [23]. These advances depend on large labeled datasets [1, 27] for training. Now, with our goal to achieve similar capabilities of detection and pose estimation capability in pigs, we need tools to build a similar pig pose database. This paper addresses the question of how to construct a labeling tool that will enable rapid and efficient annotation of pig poses.

Here we chose to use depth cameras for pig pose estimation, rather than color cameras as in [25], since depth can provide precise $3D$ joint and body positions as pigs move. However, with a lack of texture in depth, it is challenging to hand-label depth images, and not practical for pigs to wear motion capture suits. The solution we present here we call *Depth-Infrared Image Annotation Transfer* (DIAT) and it uses markers visible in infrared images that can be semi-automatically detected and tracked, and then their positions transferred to the depth images for training. Our method enables rapid labeling of pig depth images with minimal human labor, and has enabled us to train what we believe is the first depth-based pig pose detector. The following are the key contributions:

1) An image annotation transfer method in which we

[1]Steven Yik with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA yiksteve@egr.msu.edu

[2]Madonna Benjamin is an Assistant Professor, Swine Extension Veterinarian of College of Veterinary Medicine, Michigan State University, East Lansing, MI 48824, USA gemus@msu.edu

[3]Michael Lavagnino is an Academic Specialists in Mechanical Engineering, Michigan State University, East Lansing, MI 48824, USA lavagnin@egr.msu.edu

[4]Daniel Morris is a Associate Professor in Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA dmorris@msu.edu

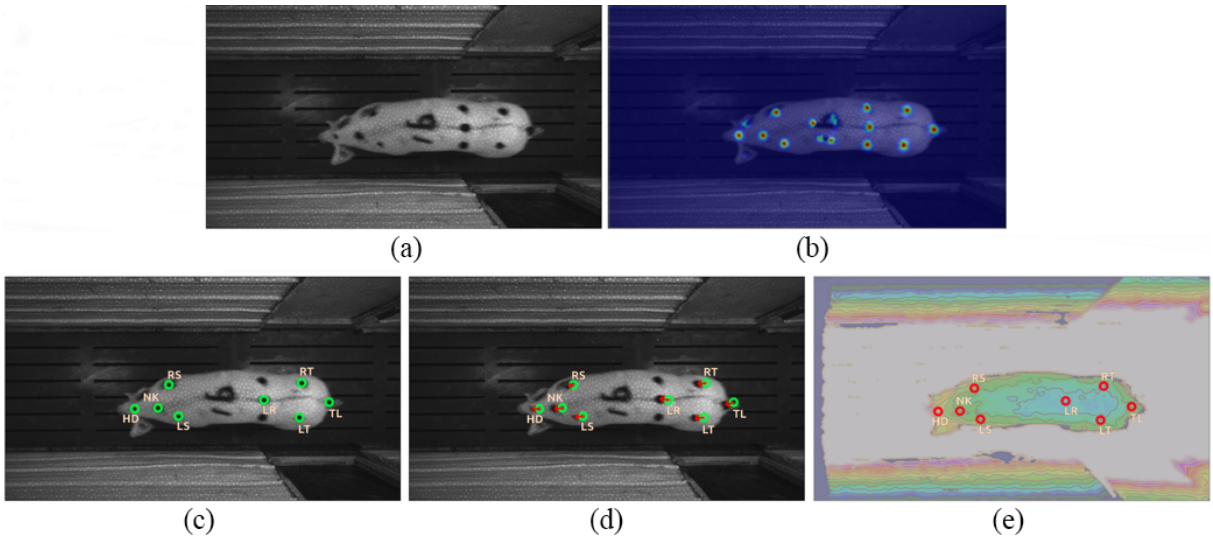

Fig. 2. An outline of the proposed *DIAT* process. (a) Infrared image of a pig showing black dots where wax crayons marked the joint locations. (b) The output of our landmark center detector shown as an overlaid heatmap. (c) Human annotated label IDs on joints used for pose. (d) Optical flow automatically associates these labels with all the other images of that pig in the video sequence. (e) Transfer of all locations and IDs from each infrared image to the corresponding depth image.

detect features in IR images and transfer them to depth images.

2) Efficient, semi-automated labeling of target joints using an IR-based landmark detector combined with optical flow for association.
3) A trained CNN for pig pose estimation from depth images.

The outline of the paper is as follows: Related work on automatic labeling are discussed in Section II. The proposed approach with high precision semi-automatic labeling is described in Section III, methodology and implementation details on how labels are obtained is described in Section IV, dataset composition and details are outlined in Section V, and experiments and results are summarized in section VI. Section VII includes our conclusions and future work intention.

II. Related Work

Pose and posture estimation have received significant attention over the last few decades. Most of the focus has been on humans with recent advances using depth cameras [16], and deep learning architectures [3, 7, 17, 23, 31]. These methods rely on large labeled datasets, and the key component that is missing for animals pose is a way to obtain similar large labeled datasets. Ideally this can be achieved with low human effort. This paper addresses this problem through our proposed DIAT technique minimizes human effort in labeling pose features.

Pose estimation for pigs in video has relied on a number of features including pig contours [14, 24], ellipse fitting [22], motion tracking and color cues [10, 11, 19] and flow-based detection [8, 9]. These methods give very coarse pig pose, and are unlikely to be as robust as modern CNN methods. Recent works [12, 36] uses CNNs to detect bounding boxes for pigs in RGB, IR, and depth images, which again is a coarse pose estimate. Using stereo depth cameras, [33] estimated detailed pig shape characteristics which involves normalizing for pose, although the method is interactive rather than automated. Recently [25] created a dataset with 4 hand-labeled landmarks on pigs and trained an hourglass CNN [23] to detect pose. This give much more detailed pose, and can be generalized to more pig environments. Our method uses a similar strategy for representing pose as landmarks, but we choose 8 landmarks corresponding to pig joints. The main difference is that we aim to detect pose on depth images, in which it is not feasible to precisely hand-label joint positions. Thus we developed a new image annotation transfer technique that enables us to semi-automatically label landmark locations with much greater precision than achievable by hand-labeling images. Also by using depth images, we are able to show quantitatively that we achieve precise estimates of pose and landmark locations.

Automated and semi-automated labeling have been widely used for medical feature annotation [4, 18, 20, 26, 30]. MRI or CT scans provide a voxel based $3D$ ground truth for a scanned image pixel-wise segmentation task. This is an example of using a secondary modality for ground truth acquisition and annotation of a primary modality for analysis. In this work, unlike these examples, IR is our secondary modality which we use to annotate depth images.

III. Proposed Approach

A. Depth Camera Motivation

It is important that sensing for PLF be robust within farm environments, not impede the motion of livestock, and not place undue burden on caregivers. At the same time, the measurement data should provide quantitative evaluation from which livestock condition and health can be inferred. Depth cameras have potential to address all of these needs. They are non-contact and can be placed out of the way above livestock routes, while simultaneously providing dense

3D shape measurements of livestock to assess body size, shape, joint and muscle locations [33], and with temporal data, may potentially observe kinematic characteristics and abnormalities including lesions, lameness, coordination and body condition[5].

What is missing from such 3D livestock data feeds, and what is needed in order to recover useful body and health metrics, is automated detection of pigs with precise estimation of their pose in each depth image. It is precisely this detection and pose estimation we seek to achieve in this paper.

B. Depth Camera Challenges

Despite holding precise 3D shape information, identifying particular joint locations on a body is challenging for a human labeler. Factors that contribute to this difficulty include: a lack of albedo texture in depth images; pig shape and size can vary significantly; and joint locations such as the hip and shoulders are often not prominent and difficult to identify through visual or palpation observation. Consequently, precise hand-labeling of pig joint locations in depth images is impractical, and is difficult even in high resolution color and IR images. Alternative high-precision methods such as motion capture suits or physical tracking markers are not feasible as pigs are inquisitive and will chew and ingest standard tracking markers.

C. Semi-Automated Annotation Labeler

The first part of our solution to these challenges with depth data is to use near-IR images that are available in many depth sensors including our RealSense cameras. This is an active camera that projects a near-IR pattern on the scene and uses a stereo IR pair to estimate depth both from the pattern and the underlying reflectance image. A consequence is that the selected IR image is pixel-wise aligned with the depth image. Other depth cameras have similar alignment between depth and IR images.

The alignment between IR and depth motivates us to propose DIAT: the process of detecting features in one modality, in this case IR, and transferring these features and their labels to the other modality, namely depth images. Once feature locations and IDs are transferred to the depth image, they can be used to directly train a depth-only pig pose estimator.

Now detecting pig joint locations in IR images with no marks is still difficult. Thus for training and evaluation we propose to mark pig joint locations with livestock-marking wax crayons that have high contrast in infrared. These crayons are food safe, easy to apply and do not impact the depth images. The eight key skeletal features we mark are located through palpation, and consist of the pigs' head (HD), neck (NK), left (LS) and right (RS) shoulders, last rib (LR), left (LT) and right (RT) hips, and tail-head (TL). To retain consistency, one swine specialist marked all pigs. Fig. 2(a) shows the placement of each mark in the infrared image and the resulting depth image.

Fig. 3. Hallway view of the system. Light blue indicates the RealSense camera field-of-view as it obtains top-down depth and IR video of pigs.

A summary of our proposed DIAT solution for building a depth-based pig pose estimator is as follows (with relevant section numbers indicated):

1) Pig joint locations are marked with wax crayons, and depth-IR video is recorded of their motion (IV-B).
2) A landmark detector finds landmark locations in all IR images, (IV-C).
3) Human labeling of landmark IDs in one IR image of each pig sequence, (IV-D).
4) Association of landmark IDs in each sequence using optical flow, (IV-D).
5) Transfer landmark IDs and locations to depth images and train depth-based pose detector, (IV-E).

These steps are illustrated in Fig. 2. The next section describes our methodology in detail.

IV. Methodology and Implementation

A. System Setup

The data collection system, which can be seen in Fig. 3, consists of a fully enclosed system on module board with external hard drives and *Intel RealSense D435* cameras with $86 \times 57\,\mathrm{deg}$ field-of-view. These cameras provide IR and depth images by projecting a dot pattern on the scene.

The system is suspended from the ceiling within a hallway of a farm. This configuration is placed such that the cameras are non-intrusive and provide full coverage of pig movement. As all pigs are moved from one area to another during normal farm operation through this hallway, this location is ideal for collecting and analyzing kinematic motion. The cameras are calibrated to operate between 1.5-2.5m above the floor with a fixed camera gain tuned to the pigs' IR reflectance.

B. Physical labeling for IR detection

The first stage of our method is to detect landmarks within the IR images. Wax crayon marks show up as relatively distinctive dark marks in IR on light skinned sows. The goal here is not to distinguish between marks, or even marks and other dark spots on the pigs, but rather to create a simple but precise mark center detector that operates only in IR. Since this is a relatively simple task, it was sufficient for a human to

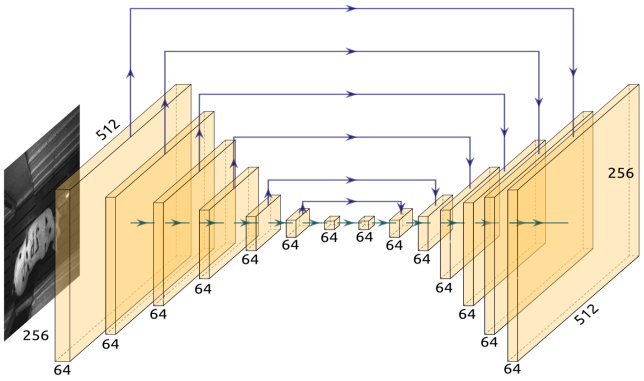

Fig. 4. Our encoder-decoder network configuration for mark detection within IR images inspired by [23, 21]. Each box indicates a stack of 3 convolutional layers with batch normalization at various sizes.

label marks in roughly 100 images of pigs in various poses and image locations. To keep the task simple we do not attempt to filter out other marks (such as numbering on the back of the pigs) and so other marks with similar appearance are given zero weight during training to avoid affecting our detector.

C. Single-Image Landmark Detection in IR

Our network shown in Fig. 4, was designed for mark detection. It uses an encoder-decoder structure consisting of a set of 3 convolution + batch norm + relu operations for each tier, each having 64 channels. This network combines features across multiple resolutions to accurately locate mark centers.

The network input is a single grayscale IR image. The label image is zero everywhere except for unit-height Gaussians with standard deviation of 6 pixels around the center of each joint location. This represents the probability of being a mark center y_i for each pixel i. The network predicts a single channel output image denoted z_i for each pixel. A weighted continuous cross-entropy loss is used for training and is specified as:

$$\mathcal{L}_{CE}(z_i, y_i) = \sum_{i=1}^{N} w_i[-y_i * z_i + log(1 + exp(z_i))] \quad (1)$$

where w_i is the pixel weight. This is minimized when the sigmoid of z_i is equal to y_i. Here the pixel weight is chosen to balance the contribution to the loss of the small number of pixels on and close to marks with the far greater number on the background. An example output, shown as a heatmap over the input image, is shown in Fig. 2(d).

D. Image Sequence Joint Labeling

After the detection of marks on a pig, identification of each joint is necessary. The approach to achieving this is to label an image where all joints are visible and to use pairwise optical flow to propagate these IDs forward and backward through a sequence. We used Deepflow [35] as it has high accuracy with large displacements which can occur when pigs run.

If only optical flow were used to propagate labels, mark locations would drift and error would accumulate. Our approach avoids this by relying on the mark detectors for location and on the optical flow only for propagating marker ID. Algorithm 1 outlines how this is done. A visualization of marking detection can be seen in Fig. 2(b).

Algorithm 1 Landmark ID labeling and association

Input: $\{I_0, \ldots, I_i, \ldots, I_n\}$ as image sequence
for every image I_i in the sequence **do**
  if $I_i == I_0$ **then**
    human assigns IDs to marks on I_i
  else
    $M \leftarrow$ do mark detection on I_i
    $F \leftarrow$ do optical flow between I_{i-1} and I_i
    for number of joints **do**
      $joint_{flow} \leftarrow F(joint)$
      search for local max on M around $joint_{flow}$
      assign max as new joint
    end for
  end if
end for
Return: joints

E. Label Transfer to Depth Images

For the task of depth image pose detection, the stacked hourglass model [23] is used as it analyzes spatial relationships on the pig. This model is similar to the architecture seen on Fig. 4, but with modifications to the skip connections, intermediate supervision, and the stacking of two hourglass networks, see Fig. 6.

Our implementation of this network is similar to the original work by Newell [23], but with the modified continuous cross-entropy loss function from Equation 1 instead of a MSE loss. As with discrete cross entropy, this loss has the advantage of maintaining gradients even when the loss is small, and so encourages better convergence. Our network consisted of 2 stacked hourglass components with 128 channels for each. The output of each hourglass is monitored and loss is evaluated to verify convergence.

V. Pig Dataset

Palpation was used to identify 8 joint locations on 158 pigs and these locations were marked with a wax crayon. Over three sessions, a dataset of 20k instances of these pigs walking or running down a hallway was collected. Both infrared and depth obtained from two RealSense™cameras were stored. These sequences are partitioned into 10% testing, 10% validation and 80% training for all neural network processing. To train the IR mark detection network, a random selection of 150 images were hand labeled, using a custom labeling program, and partitioned with the same scheme mentioned previously.

Once the IR-mark detector network is trained, it is used to find marks in the the remainder of the IR images. A human labeler selects the marks corresponding to the 8

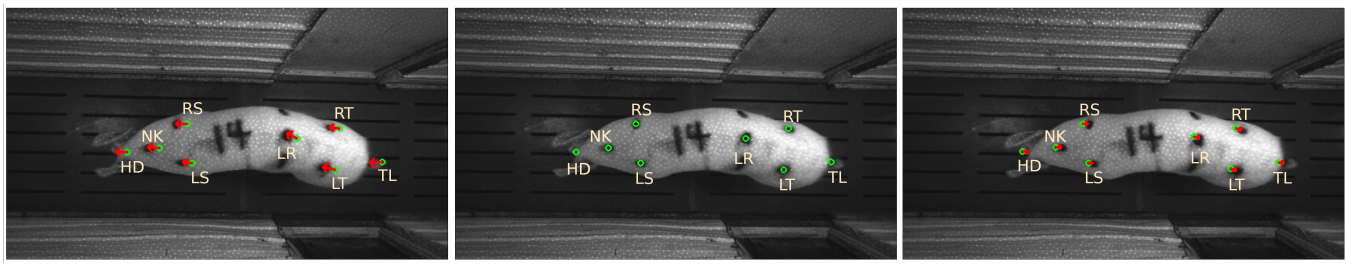

Fig. 5. **Middle:** In one image a human labeler assigns IDs to each of eight landmarks. These selected landmarks are shown in green. **Left:** Optical flow in the forward direction propagates landmark IDs from the previous image to the matched landmark location in this image. **Right:** Flow in the backward direction also propagates landmark IDs.

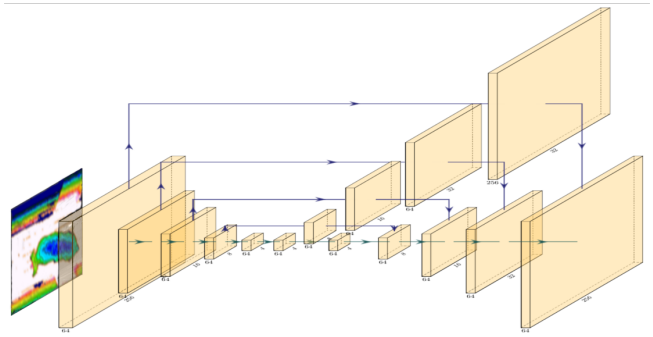

Fig. 6. The hourglass structure network for depth-based pose estimation. Each box indicates a residual module of various sizes. The completed stacked network has multiple structures appended to one to another.

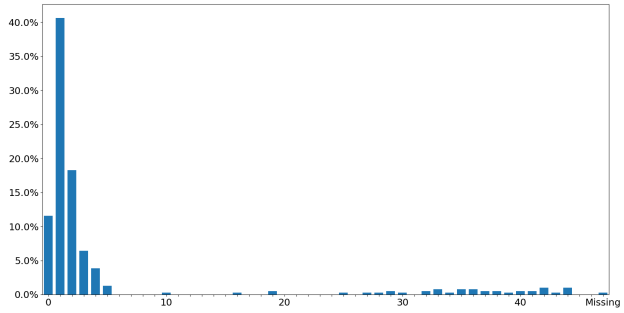

Fig. 7. Histogram of errors in estimating visible landmark centers in IR images. These locations are used to train the depth pose estimator whose differences are shown in Fig. 9 and errors in Fig. 10. The average error is 2 pixels or 0.5 cm

true landmarks and this selection is propagated through each video sequence. The previously marked ground truth locations are used to evaluate this procedure.

VI. Experiments and Results

A. Mark Detection

The output of the mark detection network can be seen in Fig. 2(*b*). The network captures all of the indicated joints with multiple false positives on other marks. These false positives are not a major concern as the fusion algorithm (Alg. 1) will filter out non-landmark detections throughout the image sequence.

Evaluation of the mark detection network is calculated based on the Euclidean pixel distance between a human labeled test set and the output of the network. The mark pixel of the network-generated mark is determined by creating a search area around the ground truth point and determining the most probable pixel. Missed detections are calculated if the maximum probability output of that region is lower than 0.3, then it is deemed a miss. Fig. 7 is a histogram of the distance error in pixels and the relative frequency of occurrence. The results show that the detection error is within a tolerance of 2 pixels with 95% confidence region for a 480x848 pixel resolution image. The missed detection rate is $< 3\%$.

B. Mark to Joint Association

The 2D skeletal pose is obtained by associating joints with detected marks in the IR images (such as those in Fig. 7). A human labeler does this in one image, and then these label IDs are propagated forward and backward in time for the full observation of a pig. There is no accumulated drift from flow as marks are detected in each image, and flow only performs association. However, there are times when association fails which usually correspond to rapid pig motion or an occlusion of a mark. These instances are automatically detected when association fails to find a detected mark with a probability peak of at least 0.3 within a 6 pixel deviation where flow predicts its location. In each of those cases, the human is prompted to either correct the association or indicate that it is occluded. Table I gives quantitative metrics on the number of human interventions during labeling showing that less than 1.3% of images require human labeling, confirmation or correction.

TABLE I

Association results for landmarks in IR images. After human assignment of IDs, only 1.3% of images need human attention.

Number of Pig Traversal Sequences	158
Average number of Images per Sequence	126
Average number of Interventions per Sequence (excluding initial labeling)	1.5
Success Rate for Automated Association	98.7%

C. Pose Detection for Depth Images

A sigmoid function is applied to the output of our pose-detector network, and results are shown in Fig. 8. Each

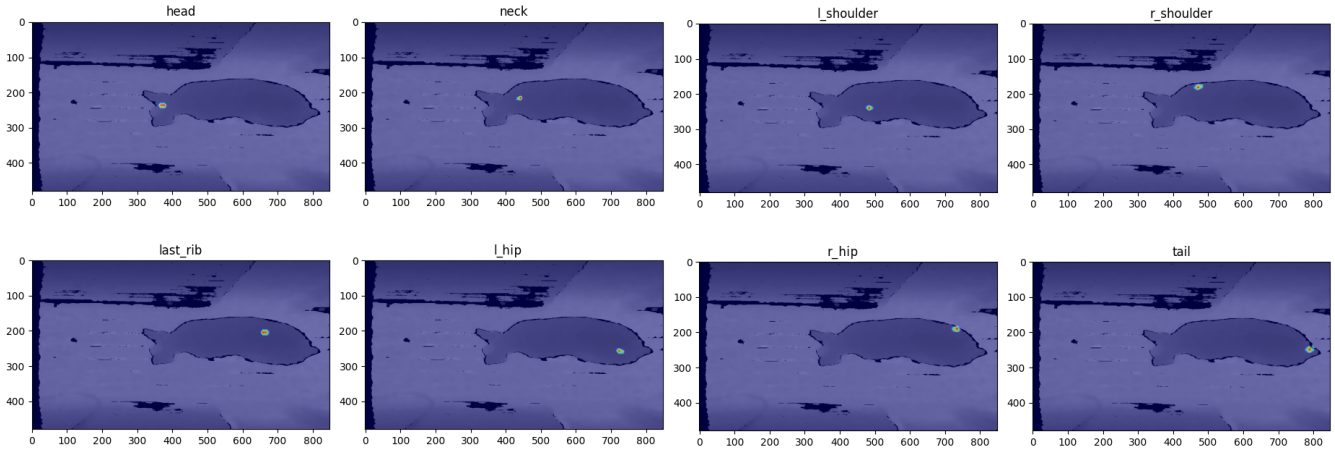

Fig. 8. Independent joint segmentation based on depth images of pigs. Each image is overlaid with the original depth image for reference. Probability of detections are denoted as 1 being red and 0 being blue.

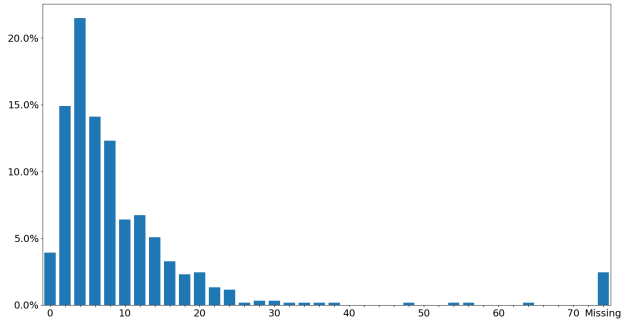

Fig. 9. Histogram of differences between IR landmark detections and depth landmark detections. One pixel is approximately 0.235 cm on the pigs' top. The average error is 8.7 pixels (2.1cm) with a 2.4% miss rate.

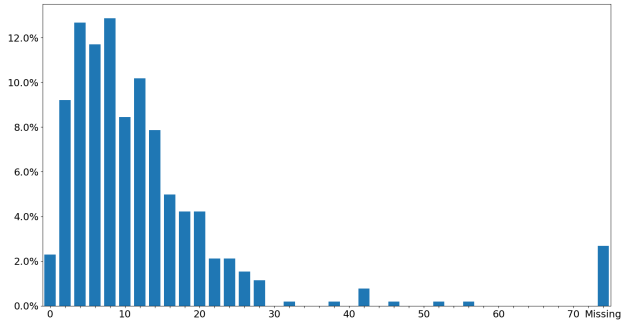

Fig. 10. Histogram of predicted joint location error from depth-images on the test partition using hand-labeled ground truth. The average error is 11.2 pixels (2.6cm) with a 2.6% miss rate.

channel represents the spatial density for a particular joint, and we select the peak as the joint's location. Together these eight locations specify a full 3D skeletal pose estimate of a pig.

These estimated joint locations can be evaluated by comparing with the human-specified mark centers in the IR images and generated labels. Histograms of these errors are shown in Figs. 9 and 10. These show an average difference of 8.7 pixels (or 2.1cm) from the detector-estimated landmark centers, and an average error of 11.2 pixels (or 2.6cm) compared to human labeled landmark centers.

Network evaluation for individual joints is made by aligning the sow horizontally using a the angle from last rib to tail as a reference. Then, a vector is made between the ground truth and the predicted joint coordinate. These values are used to calculate the covariance matrix to find axial and lateral deviation with respect to the pig's pose. Fig. 11 shows the 99.7% confidence region represented as an ellipse of each joint. Results show that there is little variation on the head joint and the most at the last rib location. This is due to a lot of contextual information around the head region and little for the last rib position. Table II show the deviation values along each axis.

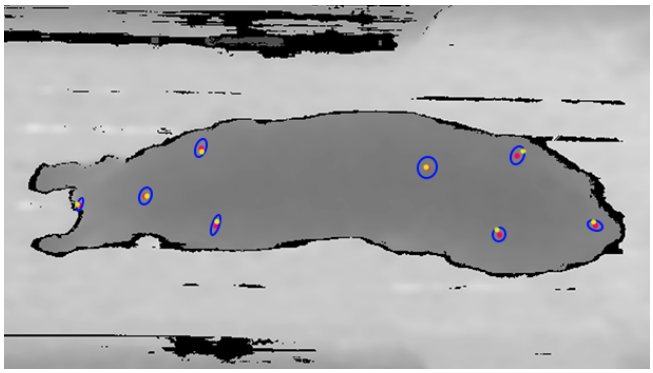

Fig. 11. Sow depth image with the ground truth joint locations (red), predicted joint locations (yellow) and the covariance ellipses (blue) at 3 standard deviations (99.7% confidence region).

VII. Conclusions

This paper proposed Depth-Infrared Image Annotation Transfer (DIAT), a novel method for annotating target pose in depth images. It uses marks, made on a target and visible in a corresponding IR image but not in the depth image. An automated IR image landmark detector is trained to

TABLE II

PREDICTED JOINT DEVIATION ALONG THE AXIAL AND LATERAL AXIS OF THE SOW FOR A 99.7% CONFIDENCE INTERVAL.

	Axial Deviation ($\sigma = 3$)		**Lateral Deviation** ($\sigma = 3$)	
Joint Name	px	cm	px	cm
Head	6.7	1.6	14.6	3.4
Neck	13.2	3.1	18.4	4.3
L shoulder	8.5	2.0	22.4	5.3
R shoulder	11.6	2.7	19.6	4.6
Last rib	19.7	4.6	21.9	5.2
L hip	13.9	3.3	14.9	3.5
R hip	13.5	3.2	19.5	4.6
Tail	15.6	3.7	10.7	2.5

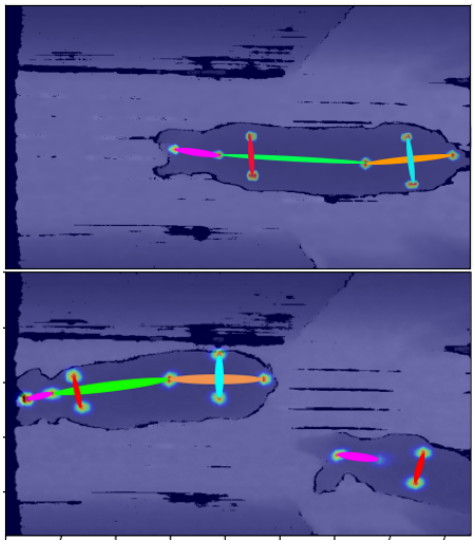

Fig. 12. Sample pig pose estimates in depth images.

precisely locate these landmarks. To label a target animal pose, a human is only required to identify each landmark in a single frame, and association with landmark locations in nearby frames is automatically performed using optical flow. Transferring these poses from IR to depth then enables training of a CNN to infer pig pose from a single depth image.

Annotation transfer of IR detections to depth is efficient, requiring very little human effort to obtain a large annotated pose dataset for depth images. Our quantitative results demonstrate that the resulting pose detector achieves an average of 11.2 pixel (2.6cm) accuracy with a 2.6% miss rate for pig poses from single depth images.

Current limitations of this method include incorrect feature association when optical flow fails for large motions. Also, while occlusion of landmarks can often be detected, it requires confirmation by the human labeler. In the future we intend to resolve these and make DIAT fully automated with no need for a human to identify the landmarks.

ACKNOWLEDGEMENTS

We thank our funding sponsors Michigan Animal Agriculture Alliance and Michigan Translational Research and Commercialization. We also thank Valley View Farms for their support in development of our system.

REFERENCES

[1] Mykhaylo Andriluka et al. “2d human pose estimation: New benchmark and state of the art analysis”. In: *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*. 2014, pp. 3686–3693.

[2] Daniel Berckmans. “Precision livestock farming technologies for welfare management in intensive livestock systems”. In: *Scientific and Technical Review of the Office International des Epizooties* 33.1 (2014), pp. 189–196.

[3] Zhe Cao et al. “OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields”. In: *arXiv preprint arXiv:1812.08008* (2018).

[4] Muyuan Chen et al. “Convolutional neural networks for automated annotation of cellular cryo-electron tomograms”. In: *Nature methods* 14.10 (2017), p. 983.

[5] Isabella Cardoso Ferreira da Silva Condotta. “Depth images\’processing to improve the performance of sows through early detection of lameness and changes in body condition score”. PhD thesis. Universidade de São Paulo, 2019.

[6] RB d’Eath. “Repeated locomotion scoring of a sow herd to measure lameness: consistency over time, the effect of sow characteristics and inter-observer reliability”. In: *Animal Welfare* (2012).

[7] Hao-Shu Fang et al. “Rmpe: Regional multi-person pose estimation”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 2334–2343.

[8] Ruta Gronskyte et al. “Monitoring pig movement at the slaughterhouse using optical flow and modified angular histograms”. In: *Biosystems Engineering* 141 (2016), pp. 19–30. ISSN: 15375110. DOI: 10.1016/j.biosystemseng.2015.10.002. URL: http://dx.doi.org/10.1016/j.biosystemseng.2015.10.002.

[9] Ruta Gronskyte et al. “Pig herd monitoring and undesirable tripping and stepping prevention”. In: *Computers and Electronics in Agriculture* 119 (2015), pp. 51–60. ISSN: 01681699. DOI: 10.1016/j.compag.2015.09.021.

[10] Yi zheng Guo et al. “Multi-object extraction from topview group-housed pig images based on adaptive partitioning and multilevel thresholding segmentation”. In: *Biosystems Engineering* 135 (2015), pp. 54–60. ISSN: 15375110. DOI: 10.1016/j.biosystemseng.2015.05.001. URL: http://dx.doi.org/10.1016/j.biosystemseng.2015.05.001.

[11] Yizheng Guo et al. “Foreground detection of group-housed pigs based on the combination of Mixture of Gaussians using prediction mechanism and threshold segmentation”. In: *Biosystems Engineering* 125 (2014), pp. 98–104. ISSN: 15375110. DOI: 10.1016/j.biosystemseng.2014.07.002.

URL: http://dx.doi.org/10.1016/j.biosystemseng.2014.07.002.

[12] Miso Ju et al. "A kinect-based segmentation of touching-pigs for real-time monitoring". In: *Sensors (Switzerland)* 18.6 (2018). ISSN: 14248220. DOI: 10.3390/s18061746.

[13] Mohammad Amin Kashiha et al. "Automatic monitoring of pig locomotion using image analysis". In: *Livestock Science* 159 (2014), pp. 141–148.

[14] Mohammadamin Kashiha et al. "The automatic monitoring of pigs water use by cameras". In: *Computers and Electronics in Agriculture* 90 (2013), pp. 164–169. ISSN: 01681699. DOI: 10.1016/j.compag.2012.09.015. URL: http://dx.doi.org/10.1016/j.compag.2012.09.015.

[15] Mark Knauer and D.J. Baitinger. "The sow body condition caliper". In: *Applied Engineering in Agriculture* 31 (Mar. 2015), pp. 175–178. DOI: 10.13031/aea.31.10632.

[16] Tsung-Yi Lin et al. "Microsoft coco: Common objects in context". In: *European conference on computer vision*. Springer. 2014, pp. 740–755.

[17] Alexander Mathis et al. "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning". In: *Nature neuroscience* 21.9 (2018), p. 1281.

[18] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation". In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE. 2016, pp. 565–571.

[19] Mateusz Mittek et al. "Health monitoring of group-housed pigs using depth-enabled multi-object tracking". In: *Proceedings of Int Conf Pattern Recognit, Workshop on Visual observation and analysis of Vertebrate And Insect Behavior*. 2016.

[20] Mehdi Moradi et al. "A cross-modality neural network transform for semi-automatic medical image annotation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 300–307.

[21] D. Morris. "A pyramid CNN for dense-leaves segmentation". In: *Proceedings - 2018 15th Conference on Computer and Robot Vision, CRV 2018*. 2018. ISBN: 9781538664810. DOI: 10.1109/CRV.2018.00041.

[22] Abozar Nasirahmadi et al. "Using machine vision for investigation of changes in pig group lying patterns". In: *Computers and Electronics in Agriculture* 119 (2015), pp. 184–190. ISSN: 01681699. DOI: 10.1016/j.compag.2015.10.023.

[23] Alejandro Newell, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation". In: *European conference on computer vision*. Springer. 2016, pp. 483–499.

[24] Andrea Pezzuolo et al. "On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera". In: *Computers and Electronics in Agriculture* 148 (2018), pp. 29–36.

[25] Eric T Psota et al. "Multi-pig part detection and association with a fully-convolutional network". In: *Sensors* 19.4 (2019), p. 852.

[26] Martin Rajchl et al. "Deepcut: Object segmentation from bounding box annotations using convolutional neural networks". In: *IEEE transactions on medical imaging* 36.2 (2016), pp. 674–683.

[27] Ben Sapp and Ben Taskar. "Modec: Multimodal decomposable models for human pose estimation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 3674–3681.

[28] Jamie Shotton et al. "Real-Time Human Pose Recognition in Parts from a Single Depth Image". In: *CVPR*. Best Paper Award. IEEE, 2011. URL: https://www.microsoft.com/en-us/research/publication/real-time-human-pose-recognition-in-parts-from-a-single-depth-image/.

[29] Sophia Stavrakakis et al. "Validity of the Microsoft Kinect sensor for assessment of normal walking patterns in pigs". In: *Computers and Electronics in Agriculture* 117 (2015), pp. 1–7.

[30] Ahmed Taha et al. "Kid-net: convolution networks for kidney vessels segmentation from ct-volumes". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 463–471.

[31] Alexander Toshev and Christian Szegedy. "Deeppose: Human pose estimation via deep neural networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 1653–1660.

[32] Erik Vranken and Dries Berckmans. "Precision livestock farming for pigs". In: *Animal Frontiers* 7.1 (2017), pp. 32–37.

[33] Ke Wang et al. "A portable and automatic Xtion-based measurement system for pig body size". In: *Computers and Electronics in Agriculture* 148.September 2017 (2018), pp. 291–298. ISSN: 01681699. DOI: 10.1016/j.compag.2018.03.018. URL: https://doi.org/10.1016/j.compag.2018.03.018.

[34] Christopher M Wathes et al. "Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall?" In: *Computers and electronics in agriculture* 64.1 (2008), pp. 2–10.

[35] Philippe Weinzaepfel et al. "DeepFlow: Large displacement optical flow with deep matching". In: *Proceedings of the IEEE international conference on computer vision*. 2013, pp. 1385–1392.

[36] Lei Zhang et al. "Automatic individual pig detection and tracking in surveillance videos". In: *arXiv preprint arXiv:1812.04901* (2018).