OpenCowID: Zero-Shot Visual Identification of Dairy Cows

Omkar Prabhune and Younghyun Kim
Purdue University
West Lafayette, IN, USA
{oprabhun, younghyun}@purdue.edu

Abstract

Accurate identification of individual cows is essential to precision dairy farming. While computer vision offers a non-invasive alternative to ear tags and RFID systems, its practical deployment remains limited by the need for zero-shot identification in dynamic herds where test identities are unseen during training. In this work, we propose OpenCowID*, a unified framework that addresses this challenge. First, we introduce a stochastic cow coat synthesis pipeline that efficiently generates large-scale, diverse images. Second, using the generated large-scale high-quality data, we present a centroid-guided feature learning strategy that forms a well-structured embedding space using virtual class centroids, enabling generalization to unseen identities.* OpenCowID *achieves state-of-the-art zero-shot and open-set identification on real-world cow benchmarks, without requiring any real labeled training data. This work contributes to the advancement of automated livestock monitoring, enabling robust, non-invasive identification. Code is available at* https://github.com/neis-lab/OpenCowID.

1. Introduction

Recent AI advances are revolutionizing dairy farming, where computer vision enables herd management and milk quality control, improving both efficiency and animal welfare [6, 12, 14].

Just as personalized healthcare begins with identifying individuals, data-driven decision-making in precision dairy farming starts with accurately recognizing each cow, which is a crucial first step toward smarter herd management. In particular, computer vision-based identification of dairy cows has gained attention as an efficient and non-invasive alternative to ear tags or RFID systems. For instance, in Holstein-Friesian cows, which is the most widely bred dairy cattle globally [5], identification is typically done by their distinctive black-and-white coat patterns, which serve as a primary biometric feature [22]. While such patterns make computer vision-based identification promising, practical deployment remains limited due to the following key challenges.

Most of the existing approaches [2–5, 11, 19, 25, 28, 30, 33, 38, 40] operate under the *closed-set* assumption, where all test identities are assumed to be present during training. However, this assumption significantly limits their practical utility in real-world scenarios. In commercial dairy farms, herd composition is highly dynamic as farmers routinely cull and introduce animals to optimize production, animal health, and farm capacity [27, 34]. The limitations of closed-set identification become apparent when encountering unknown classes during testing, often requiring retraining or modifying the model architecture, making closed-set identification impractical. This motivates the need for *open-set* identification, where some test samples have unknown identities, and *zero-shot* (disjoint-set) identification, where train and test sets share no overlapping identities.

Collecting sufficient samples per identity is especially important for training standard recognition models, and even more so for zero-shot identification methods, to enable generalization beyond the training identities. However, constructing large-scale, high-quality cow identity datasets with manual annotations is both costly and labor-intensive, and no existing dataset provides the necessary scale for this task. As a result, no existing cow identification methods are practically applicable to open-set or zero-shot settings.

In this work, we introduce OpenCowID, a novel framework designed to address both of these challenges through a unified solution.

Key contributions of this paper are as follows:

- We propose *stochastic cow coat synthesis*—an efficient image generation pipeline that synthesizes large-scale cow coat patterns tailored for biometric identification. This lightweight synthesis strategy effectively bridges the data gap and serves as an alternative to computationally heavy generative AI models, enabling scalable zero-shot identification in open-world scenarios.
- We introduce a novel *centroid-guided feature learning framework* that constructs a well-structured hyperspher-

ical embedding space through explicit centroid optimization, which outperforms the existing learning frameworks that depend on hard sample mining and implicit class separation.
- We integrate the proposed synthetic data generation pipeline and novel learning strategy into a unified framework, OPENCOWID, and demonstrate its effectiveness by achieving state-of-the-art performance on both open-set and zero-shot settings. The framework achieves cross-domain generalization and scalability without using any real labeled data.

2. Related Work

2.1. Computer vision-based cow identification

Individual cow identification based on computer vision aims to recognize visual cues in cow images to distinguish individuals within a herd. Among various methods, coat pattern-based identification [2–5, 11, 19, 25, 28, 30, 33, 38, 40] is the most common. These methods rely on RGB images of the animal's back or sides captured from a distance, making them suitable for non-intrusive, remote identification. However, these approaches assume a closed-set identification setting.

Other naturally occurring traits, such as facial features [7], muzzle patterns [23], and retinal patterns [1], have also been explored. However, these methods not only assume closed-set identification but also require highly constrained or specialized imaging conditions that are impractical for real-world commercial farms, particularly for moving cows. Furthermore, all of these approaches rely on well-annotated training data for individual cows, which is scarce and difficult to obtain at scale.

2.2. Open-set and zero-shot identification approaches

To address the limitations of closed-set models, recent methods have explored open-set identification. In [31, 32], semantic segmentation and keypoint detection are used to align cow images to a predefined template, followed by handcrafted matching with reference cows. While effective in constrained scenarios, these methods rely on accurate keypoint detection and handcrafted pipelines.

In [36], cow identification is done through ear tag detection and text recognition, but is severely limited by camera placement and tag visibility, which are often unmet constraints in real-world barns.

Inspired by human re-identification using contrastive learning [13, 26, 35], deep metric learning has been extended to cows. In [5], ResNet-50 is trained using reciprocal triplet and softmax loss to map images into a feature space, enabling clustering of both known and unknown cows. In [11], ArcFace loss is used to train a vision transformer-based foundation model for zero-shot cow identification. While this allows for open-set matching, it still requires real annotated images to extend the model to unseen identities. In addition, triplet-based training relies on hard negative mining, which can be challenging to scale and tune effectively [17, 20, 35].

2.3. Synthetic data generation for identification tasks

Recent work has explored synthetic data generation for identity-aware recognition tasks to address the issue of data scarcity. For instance, [8] uses a diffusion-based framework that synthesizes identity-conditioned human face images. In the animal domain, [29] developed finetuning method for diffusion models for identity-preserving animal image generation. These methods are resource-intensive, require large-scale annotated data, and depend on prompt engineering or identity conditioning to produce identity-aware images.

3. The OPENCOWID Framework

We introduce OPENCOWID, a novel framework for open-set and zero-shot cow identification. It comprises three key components as shown in Fig. 1.

- **Step 1: Stochastic cow coat synthesis**. This novel module automatically generates a large, diverse set of synthetic cow coat images. This effectively addresses the data scarcity problem and supports downstream training.
- **Step 2: Centroid-guided feature learning.** This is a novel two-step strategy that trains an encoder using the synthetic data: first optimizing virtual class centroids on a hypersphere, then aligning sample embeddings to them via a multi-objective loss.
- **Step 3: Zero-shot inference using *k*-NN.** Final identification is performed using k-nearest neighbors (k-NN) in the learned embedding space, enabling generalization to closed-set, open-set, and zero-shot settings without retraining.

In the following subsections, we provide a detailed description of the above components.

3.1. Stochastic cow coat synthesis

Unlike other vision tasks, no large-scale dataset exists for cow identification, and manual annotation is costly. To address limited samples per individual, our stochastic cow coat synthesis pipeline (Step 1 in Fig. 1) generates diverse large-scale training data. To synthesize high-diversity cow coat patterns, we design a lightweight, fully automated image generation pipeline. It simulates visually realistic cow coat patterns using a four-stage process: random noise initialization, spatial smoothing, binary thresholding, and photorealistic postprocessing. This section describes each step

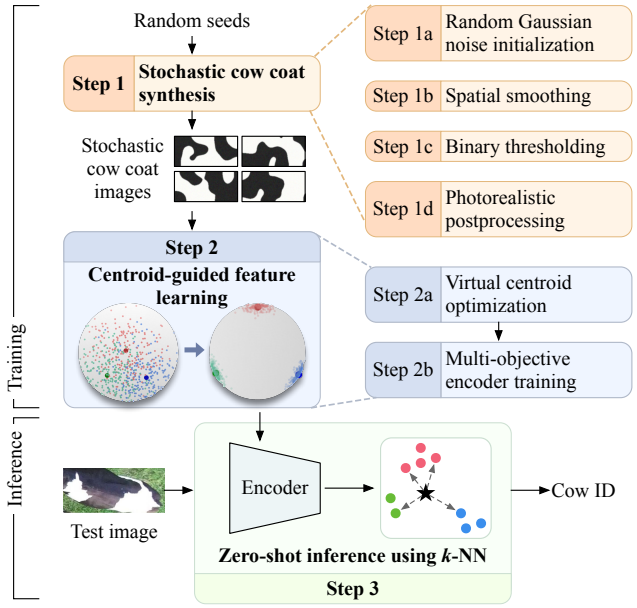

Figure 1. Overview of the OPENCOWID framework. The proposed pipeline consists of three main components: stochastic cow coat synthesis, centroid-guided feature learning, and zero-shot inference.

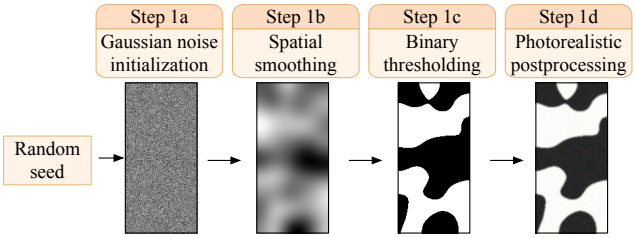

Figure 2. Example of image generation by the stochastic cow coat synthesis (Step 1). From a random seed, an image with Gaussian noise is generated. It is then smoothed, thresholded, and post-processed to simulate realistic fur color and texture.

in detail, illustrated with corresponding intermediate outputs in Fig. 2.

First, we begin by generating a 2D matrix of Gaussian noise (Step 1a), defined as:

$$I_{\text{noise}}(x, y) \sim \mathcal{N}(0, 1), \quad \forall (x, y), \tag{1}$$

where I_{noise} denotes the initial grayscale image and (x, y) are the pixel coordinates in the image.

The visual structure of the generated coat patterns is primarily determined by this initial Gaussian noise. By varying the seed, the pipeline produces a diverse range of distinct and reproducible coat patterns, each corresponding to a unique individual identity.

Next, to induce spatial coherence, we convolve the noise image with a Gaussian kernel in the spatial smoothing step (Step 1b):

$$I_{\text{smooth}} = G_\sigma * I_{\text{noise}}, \tag{2}$$

where G_σ is a Gaussian kernel with standard deviation σ, and $*$ denotes convolution. The smoothing operation encourages the formation of soft-edged, blob-like structures. Here, σ controls the 'blobiness' of the coat pattern generated—small values produce finer patterns (more fragmented), whereas larger values result in coarser blobs.

We binarize the smoothed image based on a threshold T that controls the white-to-black pixel ratio (Step 1c):

$$I_{\text{binary}}(x, y) = \begin{cases} 1 & \text{if } I_{\text{smooth}}(x, y) > T, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

This produces a binary segmentation that approximates the natural coat pattern of cows. Lower values of T yield more black regions, while higher values produce more white regions in the image.

Finally, the binary image is colorized using a realistic RGB palette (Step 1d). White and black regions are perturbed with controlled noise to simulate cow's fur texture:

$$I_{\text{RGB}}[w] = \text{clip}(C_{\text{white}} + \epsilon_w, 0, 255), \tag{4}$$

$$I_{\text{RGB}}[b] = \text{clip}(C_{\text{black}} + \epsilon_b, 0, 255), \tag{5}$$

where C_{white} and C_{black} are white and black pixels, respectively, in I_{binary}, and ϵ_w, ϵ_b are random pixel-level perturbations. Furthermore, a light Gaussian blur is applied to soften transitions, simulating cow's natural hair blending.

Additionally, we apply data augmentation techniques to each of the synthesized cow coat images to expand the training dataset, including affine and perspective transforms, color jitter, and Gaussian blur to enhance robustness and mitigate overfitting of the model during training. The resulting large-scale and diverse training set is then used to drive the proposed feature learning framework, described in the next section.

3.2. Centroid-guided feature learning

In the following centroid-guided learning framework, Step 2 in Fig. 1, the generated synthetic images are processed using an encoder that maps the input images to a hyperspherical embedding space.

The proposed learning framework trains the encoder such that it constructs an embedding space where each cow ID forms distinct and well-separated clusters. The learning process is driven by two key objectives: 1) maximizing the spatial separation between class centroids in the embedding space, and 2) encouraging the formation of compact clusters for each class.

We describe the two steps of the centroid-guided learning framework, virtual centroid optimization and multi-objective encoder training, in the following subsections.

3.2.1. Virtual centroid optimization

To enable identity-aware representation learning, we begin by defining a set of optimal target positions, termed as *virtual centroids*, for each identity class in the hyperspherical embedding space. These centroids act as target embeddings that guide the learning process. The goal of this step is to compute these virtual centroids such that they are maximally separated from each other on the unit hypersphere, thereby encouraging high inter-class discriminability.

We first compute initial centroids for each class using the output embeddings from the untrained encoder. For a class i, let $(e_1, e_2, \ldots, e_k)$ be the set of k sample embeddings belonging to that class. The initial centroid $\mathcal{C}_{\text{current}}(i)$ is defined as the normalized mean of the embeddings:

$$\mathcal{C}_{\text{current}}(i) = \text{Normalize}\left(\frac{1}{k}\sum_{j=1}^{k} e_j\right), \tag{6}$$

where normalization ensures that each centroid lies on the unit hypersphere, consistent with the L2-normalized encoder outputs.

To enforce maximal separation among class centroids, we optimize their positions by minimizing a pairwise distance loss defined as:

$$\mathcal{L}_{\text{centroid}} = \sum_{c_i \in \mathcal{C}_{\text{current}}} \sum_{c_j \in \mathcal{C}_{\text{current}}} \left(-d\left(c_i, c_j\right)\right), c_i \neq c_j, \tag{7}$$

where $d(\cdot)$ denotes the cosine or Euclidean distance between two centroids. Minimizing the negative distance effectively pushes all centroids away from one another, maximizing their spread over the hypersphere.

This optimization is performed only on the centroid vectors and does not involve any updates to the model parameters. The gradient descent is applied directly to the centroids, which are treated as free-floating learnable tensors during this step. The resulting positions constitute the final set of virtual centroids $\mathcal{C}_{\text{optimal}}$, which are used as targets during the feature learning stage that follows.

Fig. 3 illustrates this process as Step 2a, where class centroids initially cluster based on raw features, and are later redistributed to maximize inter-centroid separation. These optimal centroids, maximally spaced on the hypersphere, are frozen and act as references for the next stage of multi-objective encoder training.

3.2.2. Multi-objective encoder training

With the optimal virtual centroids $\mathcal{C}_{\text{optimal}}$ computed in Step 2a, we focus on learning an embedding function modeled as the encoder that aligns each sample's representation with its corresponding class centroid, while maintaining strong inter-class separation. To achieve this, we design a multi-objective encoder training pipeline that jointly enforces three objectives: (1) alignment of embeddings with their class-specific centroids, (2) repulsion from centroids of other classes, and (3) intra-class compactness of embeddings. Together, these objectives guide the encoder to learn a well-structured and discriminative hyperspherical embedding space. This stage updates the encoder parameters using the following loss components while keeping the optimal virtual centroids fixed.

The first component encourages each sample's embedding to move closer to its class-specific virtual centroid. This is formalized as:

$$\mathcal{L}_{\text{convergence}} = \frac{1}{|\mathcal{D}|}\sum_{e \in \mathcal{D}} d\left(e, \mathcal{C}^{e}_{\text{optimal}}\right), \tag{8}$$

where e is the embedding of a training sample, $\mathcal{D}$ is the current training batch, and $\mathcal{C}^{e}_{\text{optimal}}$ is the target virtual centroid corresponding to the class label of e.

To prevent embeddings from passing near centroids of unrelated classes during training, we introduce a repulsion term:

$$\mathcal{L}_{\text{repulsion}} = \frac{1}{|\mathcal{D}|}\sum_{e \in \mathcal{D}} \sum_{\substack{c \in \mathcal{C}_{\text{optimal}} \\ c \neq \mathcal{C}^{e}_{\text{optimal}}}} = \frac{1}{d\left(e, c\right)}, \tag{9}$$

which penalizes embeddings that approach non-corresponding centroids. Since the loss increases rapidly as distance decreases, this formulation discourages weight updates in the encoder that could lead to the mingling of embeddings between classes, thus improving the separability of the clusters in the embedding space.

To further consolidate samples of the same class, we include a compactness loss that minimizes the distance between embeddings of samples from the same class:

$$\mathcal{L}_{\text{compact}} = \frac{1}{|\mathcal{D}|}\sum_{e \in \mathcal{D}} \sum_{i \in \mathcal{D}_k} d\left(e, i\right), \tag{10}$$

where $\mathcal{D}_k \subseteq \mathcal{D}$ denotes the subset of samples in the batch that share the same class label as e.

The final loss combines all three components:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{convergence}} + \beta\mathcal{L}_{\text{repulsion}} + \mathcal{L}_{\text{compact}}, \tag{11}$$

where α and β are hyperparameters controlling the relative contribution of each objective. The combined effect of the convergence, repulsion, and compactness objectives leads to a well-structured, discriminative hyperspherical embedding space. Each class forms a compact cluster around its optimized virtual centroid, while maintaining clear separation from other clusters. This step is illustrated as Step 2b in Fig. 3.

Freezing the optimal virtual centroids obtained from the virtual centroid optimization step provides a stable target structure in the embedding space, enabling the encoder to

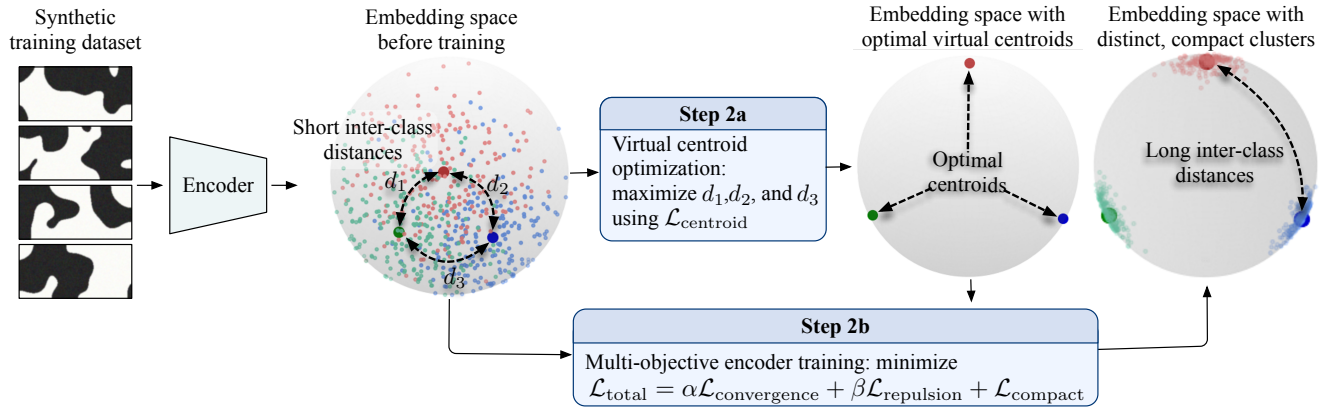

Figure 3. Proposed centroid-guided feature learning framework. Synthetic cow coat images are fed into an encoder to obtain initial hyperspherical embedding space. Virtual centroid optimization maximizes inter-class separation by adjusting virtual class centroids on the hypersphere (Step 2a). Multi-objective encoder training updates the encoder such that sample embeddings converge to their respective class centroids while maintaining inter-class separability and intra-class compactness (Step 2b).

focus solely on aligning features to these references. Since the centroids are computed to be maximally spaced on the hypersphere, they already represent the ideal inter-class separation. Allowing them to shift further would not improve class separability and could instead introduce instability during training. Moreover, if not frozen, they must be recomputed at every iteration, significantly increasing computational cost. Thus, freezing them simplifies optimization while preserving maximal inter-class distances. Once trained, the encoder is capable of generating embeddings that are identity-discriminative, even for unseen cows.

3.3. Zero-shot inference using k-NN

After training, we obtain the learned encoder $f(\cdot)$, which maps an image to a hyperspherical embedding. Given a gallery $G = \{(x_i^g, y_i)\}$ of reference images, we compute their embeddings $g_i = f(x_i^g)$. Given a query image x^q, we compute $q = f(x^q)$ and measure cosine similarity (equal to the dot product on the unit sphere) $s_i = \langle q, g_i \rangle$. We then select the k most similar neighbors $\mathcal{N}_k(q) = \text{TopK}_i\ s_i$ and predict the identity by majority vote:

$$\hat{y} = \arg\max_y \sum_{i \in \mathcal{N}_k(q)} \mathbf{1}[y_i = y]. \tag{12}$$

4. Design Merits of OpenCowID Framework

OpenCowID framework offers a comprehensive solution to the challenges of cow identification in real-world settings, notably zero-shot and open-set scenarios. In this section, we discuss the merits of the novel components in OpenCowID and how they contribute to its performance.

Comparison with existing procedural pattern synthesis methods. Edge/contrast–driven procedural textures produce crisp detail but lack direct control of connected-component size or coverage [10, 21]. Reaction–diffusion on surfaces yields spots/stripes, yet tuning PDE parameters to match Holstein-like heavy-tailed blotches is indirect [37]. Exemplar pipelines, such as graph/patch and partition-of-unity methods, require source textures, offer limited control over global pattern-level properties (e.g., black–white area ratios, blob sizes, and connected-component structure), and are not identity-aware [9, 10]. In contrast, our method explicitly targets these pattern properties, is exemplar-free and identity-aware, and reduces the synthetic-to-real gap Sec. 7 through lightweight post-processing. Two interpretable knobs, σ and T, along with the random seed directly control the generated image. Furthermore, comparing with large generative models (GANs, diffusion), we note that they are costly to train and slow to sample, and typically need sizable labeled data and careful identity conditioning/prompting—challenging at scale for diverse animal appearances [8, 16, 18, 29]. Our stochastic coat synthesis is a lightweight, exemplar-free alternative: a procedural pipeline that needs no labels, prompts, or real-image conditioning, and can generate thousands of identity-specific samples rapidly.

Our centroid-guided feature learning explicitly structures embeddings to improve encoder performance while avoiding the drawbacks of contrastive and triplet-loss frameworks [5, 13, 35], which depend on anchor–positive–negative sampling and hard mining. Instead, we adopt a two-step strategy: first optimizing virtual centroids, then clustering embeddings around them with a multi-task loss. This approach outperforms existing methods under the same setup (Sec. 6) and, combined with k-NN inference, enables zero-shot cow identification, addressing the second challenge outlined in the introduction.

5. Experimental Setup

We describe the experimental setup used to evaluate the OpenCowID framework, including details on the datasets, model architecture, training, and inference procedure.

5.1. Datasets

We use the following datasets to train and evaluate the OpenCowID framework.

To obtain training data, we use the stochastic cow coat synthesis in Sec. 3.1 to generate a large-scale, identity-rich training dataset. Specifically, we synthesize 300 unique cow identities, each representing a distinct coat pattern, followed by augmentations to synthesize 500 images per identity. This results in a total of 150,000 synthetic images for training. An additional 50 identities (i.e., 25,000 images) are generated for validation, preserving a zero-shot evaluation protocol. Samples from the synthetic training dataset are shown in Fig. 6(b).

To assess generalization in real-world conditions, we evaluate on three real-image benchmarks. **OpenCows2020** [5] comprises 4,736 images of 46 individual cows (standard test split), and **Cows2021** [15] contains 10,402 annotated images of 186 cows. In addition, we include **MultiCamCows2024 [39]**, a recent multi-camera dataset consisting of 101,329 images of 90 cows, collected in a working barn with multiple overhead cameras, introducing viewpoint, camera, and day-to-day variability.

OpenCows2020 and Cows2021 primarily provide top-down torso views under diverse indoor/outdoor lighting, whereas MultiCamCows2024 adds cross-camera evaluation with greater viewpoint/pose diversity, enabling a broader assessment of real-world generalization.

Fig. 6(a) shows some samples from these datasets. Importantly, no images from these datasets are used during training or validation for the zero-shot evaluation protocol.

5.2. Model architecture

As the encoder's backbone, we employed ResNet-50. Its output is projected into a 128-D feature space using a fully connected layer. This architectural choice of the backbone and embedding dimension follows the baseline model [5], which uses the same design. It ensures comparability with prior work and allows us to isolate the impact of our proposed framework, independent of the backbone model architecture. By maintaining this consistency, we enable a fair evaluation of the improvements introduced by our OpenCowID framework. We also perform ablation studies in Sec. 7 to compare different backbone architectures.

5.3. Training and inference procedures

For virtual centroid optimization, we initialize the encoder with ImageNet-pretrained weights and extract the initial class embeddings using the synthetic training dataset. A learning rate of 0.1 is used for the gradient descent-based optimization of centroids for 200 epochs.

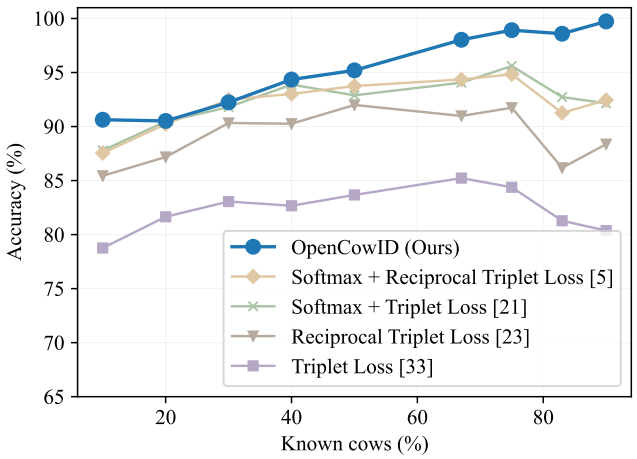

Figure 4. Open-set identification accuracy (%) of our method and existing methods [5, 24, 26, 35] across different known-unknown ratios on the OpenCows2020 dataset. Baseline results are adopted from [5].

For multi-objective encoder training, we use the Adam optimizer with a learning rate of 1×10^{-5} and a batch size of 64. The hyperparameters α and β were selected using a grid search strategy. We explored a range of values between 0.001 and 0.1 for both parameters on the validation set to determine their optimal values, which were set to 0.01 and 0.005, respectively. The model is trained for 100 epochs, with early stopping based on the validation accuracy.

For inference, k is set to 1 for k-NN, selected through cross-validation, to classify test samples in the hyperspherical embedding space. Cosine distance was used as the similarity metric, which, due to the embeddings being constrained to a hypersphere, yields identical predictions to the Euclidean distance metric.

6. Experimental Results

We evaluate OpenCowID on both open-set and zero-shot identification tasks using real-world cow datasets. These evaluations assess the framework's ability to generalize to unseen identities and compare its performance against established baseline methods.

In addition to these primary benchmarks, Sec. 7 presents ablation studies analyzing the effect of key design choices, synthetic data alternatives, and the influence of training and test set sizes on performance.

6.1. Open-set identification results

To ensure a fair comparison with existing open-set cow identification methods, we isolate the effect of our proposed centroid-guided feature learning while keeping all other experimental variables consistent with the baseline [5].

Dataset	Baseline (MegaDesc. [11])	OpenCowID (Ours)
Cows2021 [15]	73.70	**84.26±0.35** (+10.56)
OpenCows2020 [5]	94.15	**97.24±0.12** (+3.09)
MultiCamCows2024 [39]	90.53	**96.92±0.21** (+6.39)

Table 1. Zero-shot identification accuracy (%) on real datasets.

Specifically, we use the same ResNet-50 backbone and training data from the OpenCows2020 dataset, without incorporating our synthetic training data generated using the stochastic cow coat synthesis module. Following the baseline setup, the dataset is partitioned into various known/unknown identity ratios to simulate open-set identification scenarios. Only the known identities are used for training, while the test set includes both known and unknown identities.

Fig. 4 shows accuracy across different openness levels. We compare against the baseline method, which combines softmax and reciprocal triplet loss, as well as other standard contrastive learning strategies [24, 26, 35]. Our approach consistently yields higher accuracy across most openness settings, with an average improvement of 3.1%. These results highlight the effectiveness of our centroid-guided learning strategy in shaping the embedding space, since all other experimental variables are the same as the baseline. Notably, this improvement is achieved without relying on synthetic data, demonstrating better generalization to unseen identities than existing contrastive learning strategies.

6.2. Zero-shot cow identification results

For zero-shot identification, we evaluate OpenCowID on two real-world test datasets, Cows2021 [15] and OpenCows2020 [5]. We compare our method with a recent zero-shot identification method, MegaDescriptor [11], which uses Arcface loss-based metric learning [13] with Swin Transformer for feature learning.

As shown in Tab. 1, our approach outperforms the baseline by a margin of +10.56% on Cows2021, +3.09% on OpenCows2020, and +6.39% on MultiCamCows2024 in terms of accuracy. These results highlight the effectiveness of our synthetic training pipeline and the generalization capability of the centroid-guided feature learning strategy, even in the complete absence of real training identities.

7. Ablation Studies

We perform comprehensive ablation studies to isolate the effects of loss components, synthetic data generation strategy, and assess how factors such as the number of test identities, training samples, synthetic data generation parameters, domain shift, and encoder backbone influence overall performance.

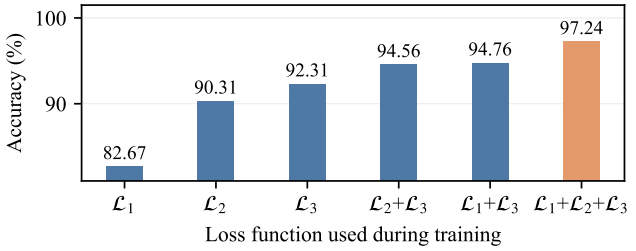

Figure 5. Effect of loss components on identification accuracy. Here, $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$, correspond to $\mathcal{L}_{\text{compact}}$, $\mathcal{L}_{\text{repulsion}}$, and $\mathcal{L}_{\text{convergence}}$, respectively.

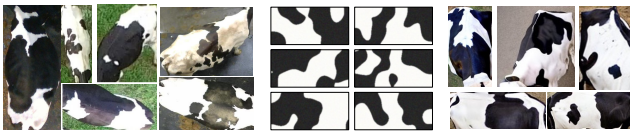

Figure 6. Samples from (a) OpenCows2020, (b) Stochastic cow coat synthesis. (c) Stable Diffusion.

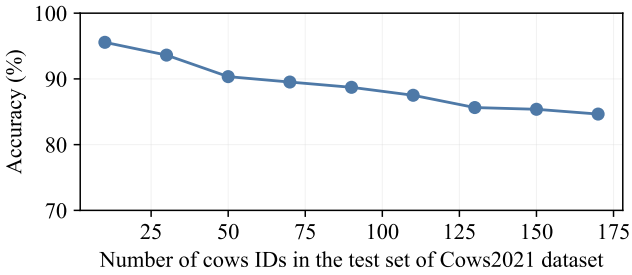

Figure 7. Effect of the number of cows in the test set on identification accuracy.

Effect of loss function components. We ablate our loss components on OpenCows2020 with 150 synthetic IDs (Fig. 5). $\mathcal{L}_{\text{compact}}$ alone performs worst, while $\mathcal{L}_{\text{convergence}}$ and $\mathcal{L}_{\text{repulsion}}$ fare better individually but improve further when combined. The full loss with all three terms achieves the best accuracy, showing that compactness, separation, and alignment provide complementary supervision essential for discriminative embeddings.

Effect of synthetic data generation method. We compare Stable Diffusion (with manual torso cropping) and our stochastic cow coat synthesis, each trained on 150 synthetic identities and tested on OpenCows2020. Our method achieves 96.77% ± 0.29 accuracy, surpassing Stable Diffusion's 96.17% ± 0.43.

Effect of test set size. We test scalability on Cows2021 [15], the largest real-world dataset with 186 cows. Trained only on synthetic data, the model is

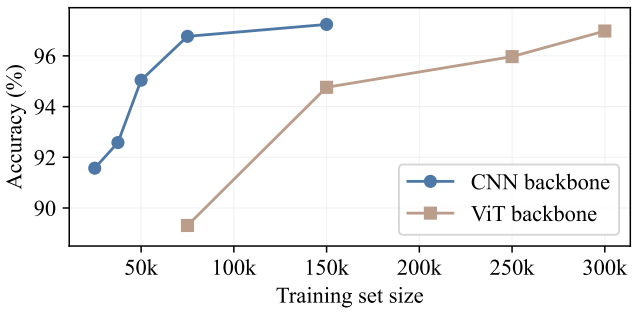

Figure 8. Effect of encoder backbone model architecture on identification accuracy, with different training set sizes.

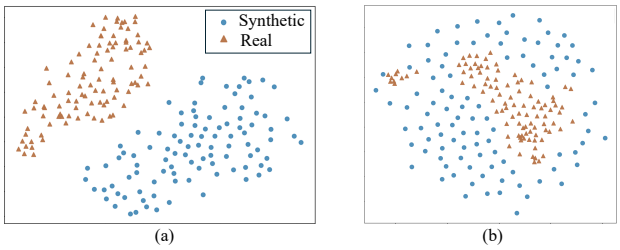

Figure 9. t-SNE of identity embeddings: (a) No augmentation vs. (b) With augmentation.

Method	FID↓	CORAL-μ↓	CORAL-Σ↓	Accuracy↑
No augmentations	1.28	0.66	0.21	80.04
With augmentations	**0.89**	**0.59**	**0.18**	**97.24**

Table 2. Domain shift analysis (synthetic vs. real), with and without postprocessing and augmentation, and effect on accuracy.

evaluated on growing subsets of unseen cows, averaging results over 10 random trials. As herd size increases, accuracy declines but remains robust, showing strong generalization to large-scale, unseen data (Fig. 7).

Effect of encoder backbone model architecture. We compare ResNet-50 and ViT (SwinV2-T) as encoder backbones. In low-data regimes, ResNet-50 outperforms ViT (91.57% vs. 89.31%), but with more synthetic data, ViT steadily improves and approaches CNN performance (Fig. 8). This shows that OpenCowID not only supports scalable learning with diverse synthetic data but also enables using data-hungry architectures like ViT, which are impractical with limited real annotations.

Synthetic-to-Real Domain Analysis and Effect of Postprocessing and Augmentations. We analyze the synthetic–to-real domain gap, quantify how our photorealistic postprocessing and augmentation stack reduces it, and assess the impact on accuracy. Fig. 9 visualizes the embedding space with t-SNE. Without postprocessing and

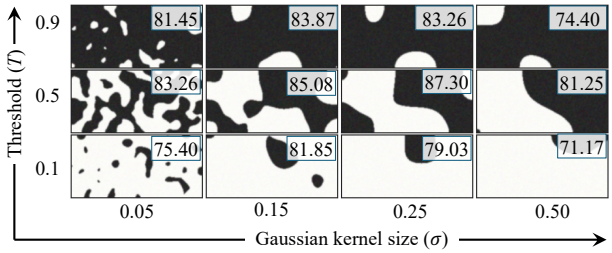

Figure 10. Effect of Gaussian kernel size (σ) and threshold (T) on the synthesized pattern for the same underlying noise. Accuracy (%) on OpenCows2020 is mentioned on the top-right corner of each configuration of (σ, T).

augmentations (Fig. 9(a)), real and synthetic embeddings form separated manifolds; realism-oriented augmentations (Fig. 9(b)) markedly reduce this separation. Quantitatively isolating the effect of our postprocessing and augmentation strategy, we observe reduced FID and CORAL scores showing reduction in the domain gap, leading to a significant increase in the accuracy of 17.2% (Tab. 2) on OpenCows2020 [5].

Effect of synthetic data generation parameters. Fig. 10 visualizes the effect of the synthesis knobs, σ and T. Keeping the training/evaluation pipeline fixed, we sweep (σ, T) and retrain for each setting. Mid-range values $\sigma \in [0.15, 0.30]$, $T \in [0.4, 0.6]$ yield coat patterns that visually resemble real data and give the best accuracy. Extremes (very small/large σ or T) produce overly fragmented or overly uniform patches that are less common, hurting generalization. Per-configuration accuracies are provided at the top-right corner of each configuration in Fig. 10.

8. Conclusion

We introduced OpenCowID, a novel framework for open-set and zero-shot dairy cow identification that addresses two critical challenges: the need for large-scale annotated data and the limitations of closed-set assumptions in dynamic real-world herds. By combining our lightweight stochastic cow coat synthesis module with a centroid-guided representation learning strategy, we enable the learning of identity-discriminative features without any real-world labels. Our unified approach generalizes well to real-world data and outperforms previous methods in both open-set and zero-shot identification settings. Beyond technical advances, OpenCowID benefits animal welfare by enabling non-invasive, accurate cow identification for improved health monitoring and herd management.

Acknowledgments

This work was supported in part by USDA National Institute of Food and Agriculture grant 2021-67021-34036 and National Science Foundation grant CNS-2112562.

References

[1] A. Allen, B. Golden, M. Taylor, D. Patterson, D. Henriksen, and R. Skuce. Evaluation of retinal imaging technology for the biometric identification of bovine animals in Northern Ireland. *Livestock Science*, 116(1–3):42–52, 2008. 2

[2] William Andrew. Visual biometric processes for collective identification of individual Friesian cattle. PhD thesis, *University of Bristol*, 2019. 1, 2

[3] William Andrew, Sion Hannuna, Neill Campbell, and Tilo Burghardt. Automatic individual Holstein Friesian cattle identification via selective local coat pattern matching in RGB-D imagery. *IEEE International Conference on Image Processing (ICIP)*, pages 484–488, 2016.

[4] William Andrew, Colin Greatwood, and Tilo Burghardt. Visual localisation and individual identification of Holstein Friesian cattle via deep learning. *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 2850–2859, 2017.

[5] William Andrew, Jing Gao, Siobhan Mullan, Neill Campbell, Andrew W. Dowsey, and Tilo Burghardt. Visual identification of individual Holstein-Friesian cattle via deep metric learning. *Computers and Electronics in Agriculture*, 185: 106133, 2021. 1, 2, 5, 6, 7, 8

[6] Ali Ismail Awad. From classical methods to animal biometrics: A review on cattle identification and tracking. *Computers and Electronics in Agriculture*, 123:423–435, 2016. 1

[7] Jayme Garcia Arnal Barbedo, Luciano Vieira Koenigkan, Thiago Teixeira Santos, and Patrícia Menezes Santos. A study on the detection of cattle in UAV images using deep learning. *Sensors*, 19(24):5436, 2019. 2

[8] Fadi Boutros, Jonas Henry Grebe, Arjan Kuijper, and Naser Damer. Idiff-Face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 19650–19661, 2023. 2, 5

[9] Jack Caron and David Mould. Partition of unity parametrics for texture synthesis. *Graphics Interface*, pages 173–179, 2013. 5

[10] Jack Caron and David Mould. Texture synthesis using label assignment over a graph. *Computers & Graphics*, 39:24–36, 2014. 5

[11] Vojtěch Čermák, Lukas Picek, Lukáš Adam, and Kostas Papafitsoros. WildlifeDatasets: An open-source toolkit for animal re-identification. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5953–5963, 2024. 1, 2, 7

[12] A. Cominotte, A. F. A. Fernandes, J. R. R. Dorea, G. J. M. Rosa, M. M. Ladeira, E. H. C. B. van Cleef, G. L. Pereira, W. A. Baldassini, and O. R. Machado Neto. Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. *Livestock Science*, 232:103904, 2020. 1

[13] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4690–4699, 2019. 2, 5, 7

[14] Rafael E. P. Ferreira, Tiago Bresolin, Guilherme J. M. Rosa, and João R. R. Dórea. Using dorsal surface for individual identification of dairy calves through 3D deep learning algorithms. *Computers and Electronics in Agriculture*, 201: 107272, 2022. 1

[15] Jing Gao, Tilo Burghardt, and Neill W. Campbell. Label a herd in minutes: Individual Holstein-Friesian cattle identification. *International Conference on Image Analysis and Processing (ICIAP)*, pages 384–396, 2022. 6, 7

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 5

[17] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 2

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:6840–6851, 2020. 5

[19] Hengqi Hu, Baisheng Dai, Weizheng Shen, Xiaoli Wei, Jian Sun, Runze Li, and Yonggen Zhang. Cow identification based on fusion of deep parts features. *Biosystems Engineering*, 192:245–256, 2020. 1, 2

[20] Mahmut Kaya and H. Bilge. Deep metric learning: A survey. *Symmetry*, 11:1066, 2019. 2

[21] Hansoo Kim, Jean-Michel Dischler, Holly Rushmeier, and Bedrich Benes. Edge-based procedural textures. *The Visual Computer*, 37(9):2595–2606, 2021. 5

[22] Hjalmar S. Kühl and Tilo Burghardt. Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in Ecology & Evolution*, 28(7):432–441, 2013. 1

[23] Santosh Kumar and Sanjay Kumar Singh. Automatic identification of cattle using muzzle point pattern: A hybrid feature extraction and classification paradigm. *Multimedia Tools and Applications*, 76:26551–26580, 2017. 2

[24] Miguel Lagunes-Fortiz, Dima Damen, and Walterio Mayol-Cuevas. Learning discriminative embeddings for object recognition on-the-fly. *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2932–2938, 2019. 6, 7

[25] Wenyong Li, Zengtao Ji, Lin Wang, Chuanheng Sun, and Xinting Yang. Automatic individual identification of Holstein dairy cows using tailhead images. *Computers and Electronics in Agriculture*, 142:622–631, 2017. 1, 2

[26] Alessandro Masullo, Tilo Burghardt, Dima Damen, Toby Perrett, and Majid Mirmehdi. Who goes there? Exploiting silhouettes and wearable signals for subject identification in multi-person environments. *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2019. 2, 6, 7

[27] Blake T. Nguyen, Kaitlyn R. Briggs, Steve Eicker, Michael Overton, and Daryl V. Nydam. Herd turnover rate reexamined: A tool for improving profitability, welfare, and sustainability. *American Journal of Veterinary Research*, 84(1), 2023. 1

[28] Fumio Okura, Saya Ikuma, Yasushi Makihara, Daigo Muramatsu, Ken Nakada, and Yasushi Yagi. RGB-D video-based individual identification of dairy cows using gait and texture analyses. *Computers and Electronics in Agriculture*, 165: 104944, 2019. 1, 2

[29] Zongming Peng, Tie Liu, Yangqianqian Chen, Yue Yang, Keren Fu, Fan Pan, and Qijun Zhao. Identity-preserving animal image generation for animal individual identification. *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 349–362, 2024. 2, 5

[30] Yongliang Qiao, Daobilige Su, He Kong, Salah Sukkarieh, Sabrina Lomax, and Cameron Clark. Individual cattle identification using a deep learning based framework. *IFAC-PapersOnLine*, 52(30):318–323, 2019. 1, 2

[31] Manu Ramesh and Amy R. Reibman. SURABHI: Self-training using rectified annotations-based hard instances for eidetic cattle recognition. *Sensors*, 24(23), 2024. 2

[32] Manu Ramesh, Amy R. Reibman, and Jacquelyn P. Boerman. Eidetic recognition of cattle using keypoint alignment. *Electronic Imaging*, 35:279–1–279–6, 2023. 2

[33] Unmesh Raskar, Omkar Prabhune, Hien Vu, and Younghyun Kim. MooBot: RAG-based video querying system for dairy cattle behavior and health insights. *Workshop on Computer Vision for Animal Behavior Tracking and Modeling (CV4Animals)*, 2025. 1, 2

[34] Josje Scheurwater, Ruurd Jorritsma, Mirjam Nielen, Hans Heesterbeek, Jan van den Broek, and Hilde Aardema. The effects of cow introductions on milk production and behaviour of the herd measured with sensors. *Journal of Dairy Research*, 88(4):374–380, 2021. 1

[35] Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A unified embedding for face recognition and clustering. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015. 2, 5, 6, 7

[36] Moniek Smink, Haotian Liu, Dörte Döpfer, and Yong Jae Lee. Computer vision on the edge: Individual cattle identification in real-time with ReadMyCow system. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 7056–7065, 2024. 2

[37] How Jiann Teo. Pattern formation and mapping for animal skin synthesis. Master's thesis, *Nanyang Technological University*, 2006. 5

[38] Hien Vu, Omkar Chandrakant Prabhune, Unmesh Raskar, Dimuth Panditharatne, Hanwook Chung, Christopher Choi, and Younghyun Kim. MmCows: A multimodal dataset for dairy cattle monitoring. *Advances in Neural Information Processing Systems (NeurIPS)*, 37:59451–59467, 2024. 1, 2

[39] Phoenix Yu, Tilo Burghardt, Andrew W. Dowsey, and Neill W. Campbell. Holstein-Friesian re-identification using multiple cameras and self-supervision on a working farm. *Computers and Electronics in Agriculture*, 237: 110568, 2025. 6, 7

[40] Sun Yukun, Huo Pengju, Wang Yujie, Cui Ziqi, Li Yang, Dai Baisheng, Li Runze, and Zhang Yonggen. Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score. *Journal of Dairy Science*, 102(11):10140–10151, 2019. 1, 2