

アピアランスとスペクトル情報に基づくペンギンの個体識別システム

昇洋太, 小篠裕子, 田中正行

概要 RGB画像を用いた外観に基づくペンギン検出と、ハイパースペクトル(HS)情報を用いたスペクトルに基づくペンギン同定を組み合わせることで、群れ画像中のペンギン個体同定システムを提案する。スペクトルに基づく同定のために、我々は多くの動物が類似した色パターンを持つという観察に触発された分類とアンサンブルアプローチを提案する。例えば、ペンギンは通常、白黒の色彩を持ち、個体間で一貫した色彩分布を持つ。提案システムの有効性は、平均平均精度(mAP)メトリックを用いて実験的に検証された。提案したアピアランスとスペクトルに基づく識別システムは、従来のアピアランスに基づく識別システムやスペクトルに基づく識別システムを大幅に上回った。

Key words: 個体識別, ハイパースペクトル画像, 画素毎分類

1. はじめに

動物の非接触画像ベースの生態学的研究は、モニタリングのための動物への接触タグ付けに通常高いコストがかかるため、非常に要求される。画像ベースのモニタリング手法は、様々な生物種を観察するために研究されてきた。これらのモニタリングの中で、動物の画像に基づく個体識別タスクは重要であり、挑戦的である。本稿では、ペンギンの個体識別を動物個体識別の1つとして取り上げる。

既存の画像ベースのペンギンの個体識別は、その腹のユニークなパターンに依存しており、通常、前方を向き、ペンギンの至近距離で撮影される⁴⁾²³⁾。しかし、この方法はペンギンの群れの画像には適さない。例えば、Fig.1に示した群れ画像では、ペンギンはすべて離れた場所から撮影されており、個体は様々な姿勢をとっているため、腹のパターンを識別することが困難である。

群れの画像からペンギンの個体識別を行うという課題に取り組む。我々のアプローチは、各ペンギンを検出し、離れた場所で撮影された画像から個体を識別する。ペンギンは、体が白と黒の羽毛で覆われ、個体の区別が付きにくい均一な外観をしている。

RGB画像に基づく物体検出手法は、

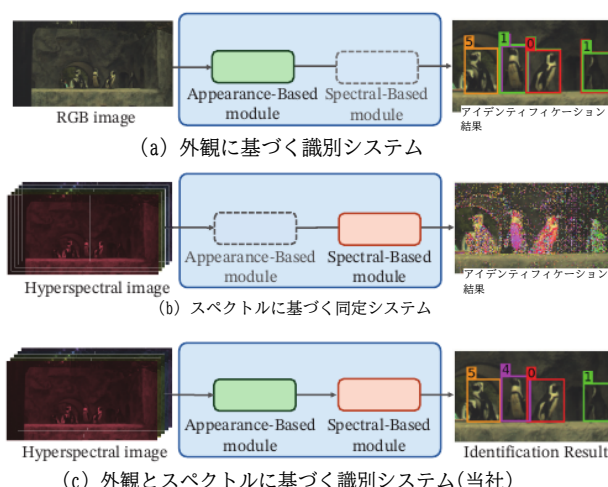


図1: 外観情報とスペクトル情報を効果的に統合する提案システム。動物の検出にはRGB画像を、個体識別にはスペクトルデータを使用する。

ディープラーニングの進歩により精度が飛躍的に向上しているにもかかわらず、外観情報のみに依存する場合、個別識別は依然として困難である。RGB画像から抽出された外観情報はペンギンの形状に着目して検出を可能にするが、ペンギンが比較的小さく見える群れ画像では、個体に関する詳細な情報を得ることは困難であり、外観のみに基づく識別を複雑にする。

我々の研究は、スペクトル情報を活用して個人識別を強化する。これまでの研究では、検出されたペンギンのピクセル単位の分類にハイパースペク

Received November 18, 2024; Revised January 23, 2025; Accepted February 12, 2025

† Tokyo Denki University

(5 Senju Asahi-cho, Adachi-ku, Tokyo, Japan)

†† Institute of Science Tokyo

(2-12-1 Ookayama, Meguro-ku, Tokyo, Japan)

【機械翻訳コンテンツの著作権について】

当サイトに掲載されている原著論文の著作権は映像情報メディア学会に帰属します。原著論文および翻訳論文の無断使用は禁止します。
引用の際には、必ず原著論文の書誌情報をご記載ください。詳細は映像情報メディア学会著作権規定をご覧ください。

Appearance-and-Spectral-Based Identification System for Penguin Individuals

Youta Noboru[†], Yuko Ozasa[†], Masayuki Tanaka^{††}

Abstract We present a system for identifying individual penguins in flock images by combining appearance-based penguin detection with RGB images and spectral-based penguin identification with hyperspectral (HS) information. For the spectral-based identification, we propose a classification and ensemble approach inspired by the observation that many animals have similar color patterns. For example, penguins typically have black and white coloration, with consistent color distributions across individuals. The proposed system's effectiveness was experimentally validated with a mean Average Precision (mAP) metric. The proposed Appearance-and-Spectral-Based identification system significantly outperformed conventional appearance-based and spectral-based systems.

Key words: Individual Identification, Hyperspectral Image, Pixel-Wise Classification

1. Introduction

Non-contact image-based ecological studies of animals are highly demanded because contact tagging animals for monitoring usually requires a high cost. Image-based monitoring methods have been studied to observe a wide range of species. Among those monitoring, the image-based individual identification task for animals is important and challenging. In this paper, we focus on the penguin's individual identification as one of animal individual identifications.

Existing image-based penguin individual identification relies on unique patterns of their bellies, typically captured in front-facing and in close range of penguins.⁴⁾²³⁾ However, this method is not suitable for images of penguin flocks. For example, in the flock image shown in Fig.1, all penguins are captured from a distance, with individuals in various postures, making it difficult to identify belly patterns.

We address the challenge of identifying individual penguins within flock images. Our approach detects each penguin and identifies individuals from images captured at a distance. Penguins have a uniform appearance with bodies covered in white and black feathers, making it difficult to distinguish individuals.

Individual identification remains challenging when re-

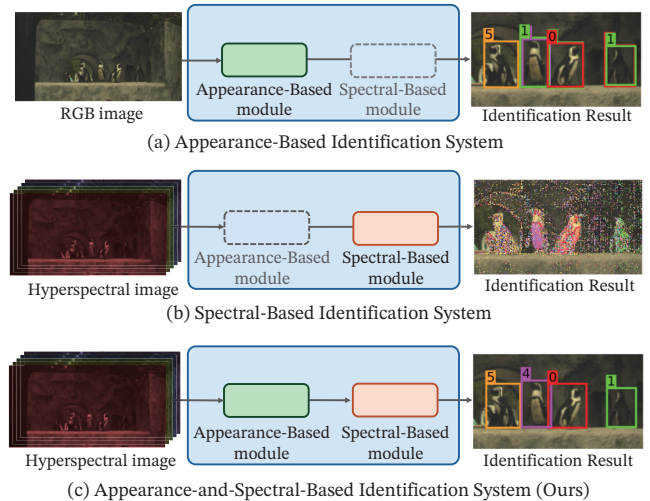


Fig. 1: Proposed system that effectively integrates appearance and spectral information. It uses RGB images for animal detection and spectral data for individual identification.

lying solely on appearance information, even though the object detection methods based on RGB images have dramatically improved in accuracy due to advances in deep learning. While appearance information extracted from RGB images enables detection by focusing on the shape of penguins, in flock images where penguins appear relatively small, obtaining detailed information about individuals is challenging, complicating identification based solely on appearance.

Our work leverages spectral information to enhance individual identification. Previous work has used Hyperspectral (HS) images for the pixel-wise classification

Received November 18, 2024; Revised January 23, 2025; Accepted February 12, 2025

[†] Tokyo Denki University
(5 Senju Asahi-cho, Adachi-ku, Tokyo, Japan)

^{††} Institute of Science Tokyo
(2-12-1 Ookayama, Meguro-ku, Tokyo, Japan)

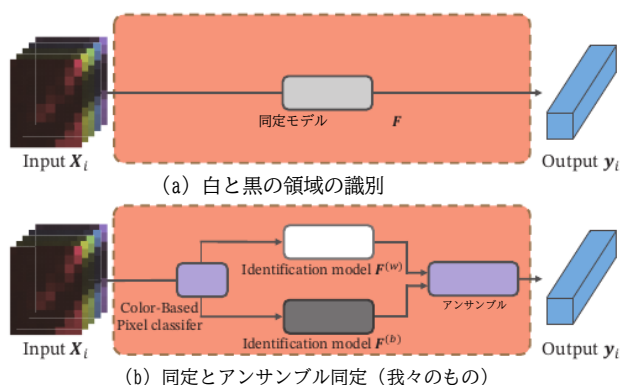


図2: ベースラインの単一モデルアプローチと比較した、我々の提案する同定とアンサンブル手法の説明図。

トル(HS)画像を使用し、個体の識別におけるスペクトル情報の有効性を実証してきた。

本論文では、図1に示すように、外観情報とスペクトル情報の両方の長所を組み合わせた個別識別システムを紹介する。具体的には、外観情報に基づくペンギン検出モジュールと、スペクトル情報に基づく個別識別モジュールを統合したシステムである。本システムは、HS画像から得られたRGB画像からペンギンを検出し、検出された領域のHSデータを用いて個体を識別する。

ペンギン個体識別モジュールでは、識別とアンサンブルの方法を提案する。図2は、2つの個別識別モデルの結果を組み合わせた提案手法と、個別識別に1つのモデルを用いたベースラインとを組み合わせた提案手法である。このアプローチは、個々の動物がしばしば似たような配色を示すという観察からヒントを得たものである。例えば、ペンギンは黒と白の2色で構成されていることが多い。そこで、まず各ペンギンの身体領域内の画素値を2つのクラスにクラスタリングし、アンサンブル法を適用して最終的な識別結果を得る。

実験では、群れ画像からペンギンの個体識別を行う際の、本システムの性能を評価した。異なる画像解像度において、我々のアプローチを外観ベースおよびスペクトルベースの手法と比較し、我々のシステムの頑健性を実証する。視覚的な比較により、困難な条件下でペンギンを正確に検出・識別する上で、本手法の有効性が浮き彫りになった。アブレーション研究により、我々の重要な構成要素である同定法、アンサンブル法、重複抑制の重要性が確認された。

我々の結果は、システムの優れた性能を実証し、ペンギン識別における白と黒の領域からのスペクトル署名の補完的な役割を強調し、このタスクのためのRGBとHSデータを組み合わせることの有効性を実証する。

2. 関連研究

2.1 カメラを用いた個体識別について

カメラを用いた個体識別は、生態学的研究において重要な課題である。シマウマやトラなどの特徴的なパターンを持つ動物は、比較的単純な同定課題である¹¹⁾。一方、ペンギンを含む多くの種は個体識別がかなり困難である。このような困難なケースでは、顔の特徴⁷⁾¹⁴⁾²¹⁾のような特定の解剖学的部位の分析に頼って識別されることが多い。例えば、パンダの顔の特徴に注目した研究では、個人識別において71.17%の精度を達成した¹⁴⁾。ペンギンでは、腹上の点の位置によって個体が識別される⁴⁾²³⁾。我々の研究は、これまでの手法が苦手としてきた主要な課題に取り組むことで、グループ画像内の動物個体識別に対する既存のアプローチを前進させるものである。動物群の画像では、異なる方向を向いている動物や、画像内の大きさが異なる動物などの問題があり、各個体の詳細な情報に基づいて、これまでの方法で個体を識別することは不可能である。

2.2 アピアランス情報に基づく動物の検出

我々のタスクは、動物検出のための先行研究とは異なる。ほとんどの研究は動物を検出し、動物の数をカウントしている¹⁸⁾が、我々は動物の数をカウントせず、個体を識別する。Wuらは衛星画像中のペンギンを検出するためにYOLO²⁵⁾を用い、また、YOLOを用いてペンギンを検出した。

2.3 スペクトル情報に基づく動物分析

ペンギン個体間のスペクトル解析に関する研究¹⁷⁾は、HSピクセルからの個体識別の実現可能性を示している。多くの研究者が、HS画像に基づいて視覚的に類似した生物のスペクトル情報を解析している。スペクトル情報は、爬虫類⁵⁾²⁰⁾²²⁾、昆虫¹⁶⁾、魚類¹⁰⁾など、種によって異なることが実証されている。

しかし、これまでの研究では、外観情報を利用しないHS画素に焦点を当てており、撮影画像から対象画素を手動で選択する必要がある。

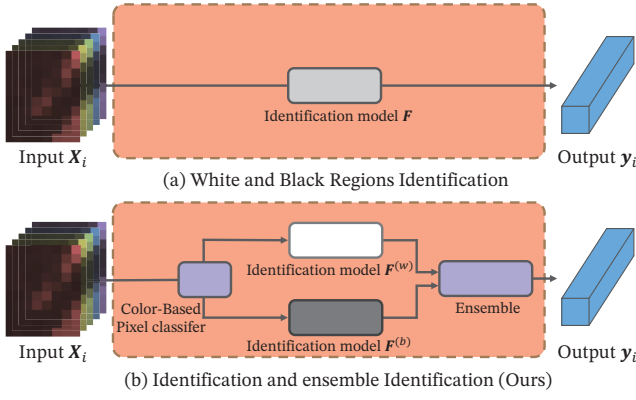


Fig. 2: Illustration of our proposed identification and ensemble method compared to the baseline single-model approach.

of detected penguins, demonstrating the effectiveness of spectral information in distinguishing individuals.

This paper presents an individual identification system that combines the strengths of both appearance and spectral information, as shown in Fig. 1. Specifically, the system integrates a penguin detection module based on appearance information and an individual identification module based on spectral information. The system detects penguins based on the RGB image derived from the HS image and then identifies individuals using the HS data of the detected regions.

In the penguin individual identification module, we propose an identification and ensemble method. Fig. 2 demonstrates our proposed method, which combines the results of two individual identification models, with the baseline, which uses a single model for individual identification. This approach is inspired by the observation that individual animals often exhibit similar color schemes. For example, penguins are predominantly composed of two colors, black and white. Therefore, we first cluster the pixel values within the body region of each penguin into two classes and then apply an ensemble method to produce the final identification result.

In the experiments, we evaluate the performance of our system in identifying individual penguins in flock images. We compare our approach with appearance-based and spectral-based methods at different image resolutions, demonstrating the robustness of our system. Visual comparisons highlight our method's effectiveness in accurately detecting and identifying penguins under challenging conditions. An ablation study confirms the importance of our key components: the identification and ensemble methods and duplicate suppression. Our results demonstrate the system's superior

performance and emphasize the complementary role of spectral signatures from white and black regions in penguin identification, demonstrating the efficacy of combining RGB and HS data for this task.

2. Related Work

2.1 Identification of Individual Animals using Cameras

Individual animal identification using cameras is a critical task in ecological research. Animals with distinctive patterns, such as zebras and tigers, present relatively straightforward identification challenges¹¹⁾. In contrast, many species, including penguins, are considerably more difficult to discriminate individuals. For these challenging cases, the identification often relies on analyzing specific anatomical regions, such as the facial characteristics⁷⁾¹⁴⁾²¹⁾. For instance, a study focusing on the facial features of pandas achieved a 71.17% accuracy in individual identification¹⁴⁾. In penguins, individuals are identified by the position of a point on the belly⁴⁾²³⁾.

Our work advances existing approaches to individual animal identification within group images by tackling key challenges that previous methods have struggled with. In images of groups of animals, problems such as animals facing in different directions, and animals of different sizes in the image, make it impossible to identify individuals using previous methods based on detailed information of each individual.

2.2 Animal Detection based on Appearance Information

Our task is different from the previous work for animal detection. Most work detects animals and counts the number of animals¹⁸⁾, but we do not count the number of animals but identify individuals. Wu *et al.* used YOLO to detect penguins in satellite images²⁵⁾ and we also used YOLO to detect penguins as they do.

2.3 Animal Analysis based on Spectral Information

Work on analyzing the spectra among individual penguins¹⁷⁾ shows the feasibility of individual identification from HS pixels. Many researchers have analyzed the spectral information of visually similar organisms based on HS images. It has been demonstrated that spectral information differs among species, for example, reptiles⁵⁾²⁰⁾²²⁾, insects¹⁶⁾, and fish¹⁰⁾.

However, previous work focuses on HS pixels without utilizing appearance information and requires manual selection of target pixels from captured images. The proposed system automatically selects pixels effective

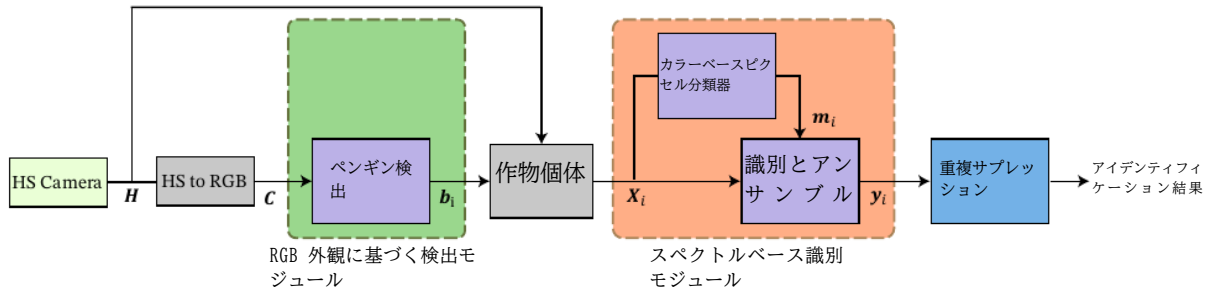


図3: ペンギンの個体識別のための提案システムの構成図。本システムには、外観の特徴に基づいてペンギンを定位する検出モジュールと、HSデータを利用してスペクトルの特徴によってペンギンの個体を識別する識別モジュールが含まれている。重複抑制処理により、各ペンギンに固有のIDを割り当て、人口密度の高い群れでも正確な識別を保証する。

提案システムは、識別に有効な画素を自動的に選択し、HS画像からペンギンの個別識別を行う。

3. 個人ペンギン識別システム

本節では、HS画像を用いたペンギンの個体識別システムについて紹介する。我々のアプローチは、HSカメラを用いたHS画像 $H \in \mathbb{R}^{H \times W \times B}$ の取得から始まり、 H 、 W 、 B はそれぞれスペクトルバンドの高さ、幅、数を表す。本システムは、外観情報とスペクトル情報の両方を活用し、ペンギンの位置特定のための検出モジュールと、個人認識のための識別モジュールを組み込んでいる。

図3は、提案システムの全体的なアーキテクチャを示したものである。以下のサブセクションでは、主要なモジュールについて説明する: 外観に基づく検出モジュール、スペクトルに基づく識別モジュール、および識別結果を洗練する重複抑制プロセス。

3.1 外観に基づく検出モジュール

外観に基づく検出モジュールは、RGB画像中の個々のペンギンの位置を特定するために設計されている。

HS画像 H は、まずCIE 1931 XYZ色マッチング関数¹³⁾を用いてRGB画像 $C \in \mathbb{R}^{H \times W \times 3}$ に変換される。ここで、3は結果のRGB画像の3つのカラーチャンネルを示す。変換されたRGB画像 C は、その後の検出とローカライズの処理に使用される。

1) ペンギン検出

RGB 画像 $C \in \mathbb{R}^{H \times W \times 3}$ は、外観に基づくペンギン検出モジュールに供給される。この外観に基づくペンギン検出モジュールにはYOLO⁸⁾を採用する。YOLOはペンギンデータセット¹⁷⁾を用いてペンギン検出用に微調整されている。

検出処理により、画像内のペンギンの位置が決定される。 P 個の検出されたペンギンのバウンディングボックスを出力し、 $b_i = [\text{left}_i, \text{top}_i, \text{right}_i, \text{bottom}_i]$ と表される。 P は検出された各ペンギンのインデックスを表し、 P は画像内で検出されたペンギンの総数を表す。

2) 作物個体

クロップ個体成分では、検出されたペンギンのバウンディングボックスに基づいて、元のHS画像 $H \in \mathbb{R}^{H \times W \times B}$ から注目領域を切り出す。この処理により、検出された P 個のペンギンのバウンディングボックスのそれぞれについて、HS画像パッチ $X_i \in \mathbb{R}^{h_i \times w_i \times B}$ が生成される。 P 、 h_i と w_i の次元は、各画像パッチの高さと幅に対応する。スケーリングファクター $r \in [0, 1]$ は、バウンディングボックスの中心付近でクロップサイズを対称に調整するために採用される。

この操作は次のように定義される:

$$X_i = \{H_{(u,v)} \mid \frac{\text{left}_i + \text{right}_i - r(\text{right}_i - \text{left}_i)}{2} < u < \frac{\text{left}_i + \text{right}_i + r(\text{right}_i - \text{left}_i)}{2}, \frac{\text{top}_i + \text{bottom}_i - r(\text{bottom}_i - \text{top}_i)}{2} < v < \frac{\text{top}_i + \text{bottom}_i + r(\text{bottom}_i - \text{top}_i)}{2}\}, \quad (1)$$

ここで、 $H_{(u,v)}$ は座標 (u, v) における H の画素を表し、 $u \in \{1, \dots, W\}$ 、 $v \in \{1, \dots, H\}$ 。

3.2 スペクトルベース識別モジュール

スペクトルベース識別モジュールは、各HS画像パッチ X_i のペンギンIDを識別する。このモジュールは、各ペンギン画素を白画素と黒画素に分類するカラーベース画素分類器から構成される。次に、白画素と黒画素をそれぞれ白画素と黒画素の識別モジュールに入力する。

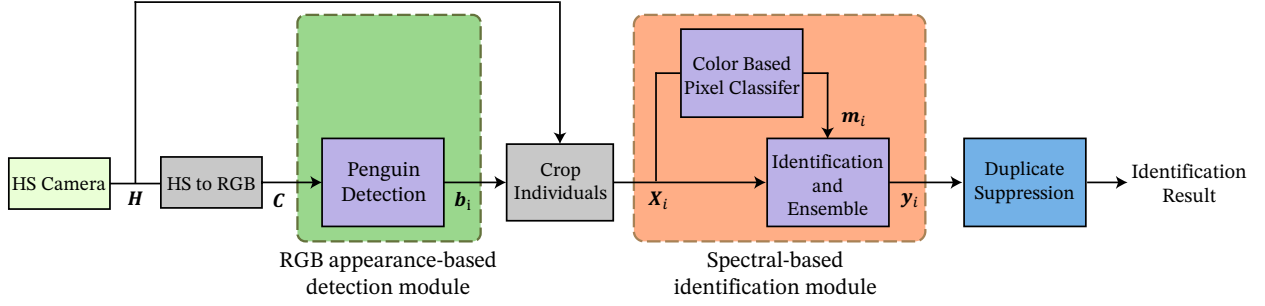


Fig. 3: Diagram of the proposed system for individual penguin identification. The system includes a detection module that localizes penguins based on appearance features, followed by an identification module utilizing HS data to distinguish individual penguins through spectral characteristics. A duplicate suppression process assigns unique identities to each penguin, ensuring accurate identification even in densely populated flocks.

for identification and performs individual penguin identification from HS images.

3. Individual Penguin Identification System

This section introduces our system for individual penguin identification using HS images. Our approach begins with the acquisition of an HS image $\mathbf{H} \in \mathbb{R}^{H \times W \times B}$ using an HS camera, where H , W , and B represent the height, width, and number of spectral bands respectively. Our system leverages both appearance and spectral information, incorporating a detection module for penguin localization and an identification module for individual recognition.

Figure 3 illustrates the overall architecture of our proposed system. The following subsections explain the key modules: the appearance-based detection module, the spectral-based identification module, and the duplicate suppression process, which refines the identification results.

3.1 Appearance-based Detection Module

An appearance-based detection module is designed to localize individual penguins in an RGB image.

An HS image $\mathbf{H}$ is first converted into an RGB image $\mathbf{C} \in \mathbb{R}^{H \times W \times 3}$ with CIE 1931 XYZ color matching functions¹³⁾. Here, the 3 indicates the three color channels of the resulting RGB image. The converted RGB image $\mathbf{C}$ is then used in the subsequent detection and localization processes.

(1) Penguin Detection

The RGB image $\mathbf{C} \in \mathbb{R}^{H \times W \times 3}$ is fed into the appearance-based penguin detection module. We adopt YOLO⁸⁾ for this appearance-based penguin detection module. The YOLO has been fine-tuned for penguin detection with the penguin dataset¹⁷⁾.

The detection process determines the locations of penguins within the image. It outputs bounding boxes for P detected penguins, represented as $\mathbf{b}_i = [\text{left}_i, \text{top}_i, \text{right}_i, \text{bottom}_i]$, where $i = \{1, \dots, P\}$ denotes the index of each detected penguin and P denotes the total number of penguins detected within the image.

(2) Crop Individuals

In the crop individuals component, we crop regions of interest from the original HS image $\mathbf{H} \in \mathbb{R}^{H \times W \times B}$ based on the detected penguin's bounding box. This process generates HS image patches $\mathbf{X}_i \in \mathbb{R}^{h_i \times w_i \times B}$ for each of the P detected penguin's bounding box, where $i = \{1, \dots, P\}$. The dimensions h_i and w_i correspond to the height and width of each image patch. A scaling factor $r \in [0, 1]$ is employed to symmetrically adjust the crop size around the center of the bounding box.

The operation is defined as:

$$\begin{aligned} \mathbf{X}_i = \{ \mathbf{H}_{(u,v)} \mid & \frac{\text{left}_i + \text{right}_i - r(\text{right}_i - \text{left}_i)}{2} \\ & < u < \frac{\text{left}_i + \text{right}_i + r(\text{right}_i - \text{left}_i)}{2}, \\ & \frac{\text{top}_i + \text{bottom}_i - r(\text{bottom}_i - \text{top}_i)}{2} \\ & < v < \frac{\text{top}_i + \text{bottom}_i + r(\text{bottom}_i - \text{top}_i)}{2} \}, \end{aligned} \quad (1)$$

where $\mathbf{H}_{(u,v)}$ represents the pixel of $\mathbf{H}$ at coordinates (u, v) , with $u \in \{1, \dots, W\}$ and $v \in \{1, \dots, H\}$.

3.2 Spectral-based Identification Module

The spectral-based identification module identifies the penguin ID for each HS image patch $\mathbf{X}_i$. This module comprises a Color-Based pixel Classifier that classifies each penguin pixel into white and black pixels. Then, white and black pixels are fed into white pixel and black pixel identification modules, respectively. Fi-

最後に、これらの識別結果をアンサンブルして、最終的な識別出力を生成する。

(1) カラーベース画素分類器

色ベース画素分類器は、ペンギンの切り出しHS画像 X_i を、そのスペクトル特性に基づいて白と黒の領域に分割する。このセグメンテーションは、その後の個体識別に不可欠であり、異なるペンギン領域の専門的な処理を可能にする。セグメンテーション結果は、2値行列 $m_i \in \mathbb{R}^{h_i \times w_i}$ として表され、これを重み付け行列と呼ぶ。図4は、ペンギン画像における白と黒の領域の分離を示す、これらの重み付け行列 m_i を可視化したものである。

ガウス混合モデル(GMM)¹⁾を採用し、画素値を2つのクラスにクラスタリングすることで、画像間の照明条件の変化に適応し、このセグメンテーションを行う。GMMは、各入力切り出し領域におけるこれらの変動を考慮するように学習される。重み付け行列の各画素 m_{ij} の分類は、以下のように決定される。

$$m_{ij} = \begin{cases} 1 & \text{if } N(x_{ij}; \mu_w, \Sigma_w) > N(x_{ij}; \mu_b, \Sigma_b) \\ 0 & \text{else} \end{cases}, \quad (2)$$

ここで、 $j = (u_i, v_i)$ $u_i \in \{1, \dots, w_i\}$, $v_i \in \{1, \dots, h_i\}$ は切り出し領域内の画素座標を表し、それぞれ幅と高さの次元を表す。 $N(-)$ はガウス確率密度関数を表し、 $\mu_w, \Sigma_w, \mu_b, \Sigma_b$ はそれぞれ白と黒のクラスの平均と共分散行列である。

どのGMMクラスが白い領域を表しているかを決定するために、2つのクラスター化されたクラスの平均画素値を比較する。平均値が高いクラスを白色クラスとし、その画素に m_{ij} の値 1 を割り当てる。逆に、平均値の低いクラス(黒い領域を表す)の画素は、 m_{ij} に 0 が割り当てられる。このアプローチにより、初期GMMクラスタリング結果に関わらず、 m_i が一貫してペンギンの白色領域に対応することが保証される。(2) 同定とアンサンブル

識別とアンサンブルコンポーネントは、個々のペンギン識別のための2つのモデルを活用する。このアプローチにより、ペンギンの体のさまざまな部位の明確なスペクトル特性を利用することができる。

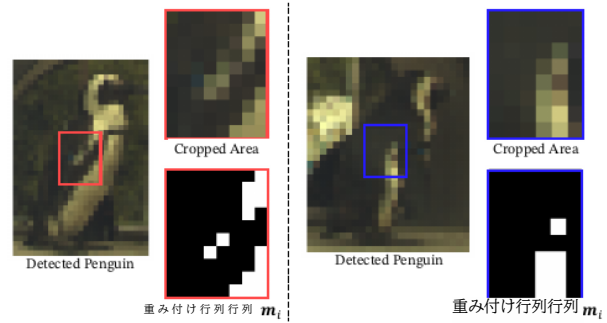


図4: ペンギン画像の重み付け行列の可視化。行列の白い部分は、ペンギンの白い部分を表す値1のピクセルに対応する。黒人の地域は、ペンギンの黒人地域に対応する値0である。

各切り出し領域 X_i に対して、2つのモデルを適用する: $F^{(w)}: \mathbb{R}^B \rightarrow \mathbb{R}^N$ と $F^{(b)}: \mathbb{R}^B \rightarrow \mathbb{R}^N$ 。これらのモデルは、それぞれ白と黒の領域に分類されたHSデータで学習される。各モデルはB次元のHS画素値を入力とし、N次元のベクトルを出力する。ここで、Nはデータセット中のペンギンIDの数に対応する。

これらの領域別モデルからの出力を統合するために、色ベース画素分類器から得られる重み付け行列 m_i を用いた重み付け平均を採用する。この方法は、ピクセル単位のセグメンテーションに基づき、白と黒の両方の領域からのスペクトル情報を結合する。各バウンディングボックスの同定信頼度は以下のように計算される:

$$y_i = \frac{1}{h_i \times w_i} \sum_{j=1} \{m_{i,j} F^{(w)}(x_{i,j}) + (1 - m_{i,j}) F^{(b)}(x_{i,j})\}, \quad (3)$$

ここで、 $y_i \in \mathbb{R}^N$ はi番目に検出されたペンギンの信頼度ベクトルであり、バウンディングボックス内のピクセル単位の信頼度スコアの加重平均として計算される。

このアンサンブルアプローチにより、各ペンギンIDの信頼度スコアが白と黒の両方の地域からの情報を反映することが保証され、個人識別精度が向上する可能性がある。

3) 識別モデルの学習

識別モデル $F^{(w)}$ と $F^{(b)}$ は、HS画像から個々の画素を処理し、対応するペンギンIDを予測するように設計されている。両モデルとも、5つの完全接続層からなる多層パーセプトロン(MLP)として実装されている。このアーキテクチャは、ペンギン識別のためのHSデータの処理に有効であることが示されている先行研究¹⁷⁾から採用した。

これらのモデルを学習するために、

nally, those identification results are ensembled to generate the final identification output.

(1) Color-Based Pixel Classifier

The Color-Based Pixel Classifier segments the cropped HS images $\mathbf{X}_i$ of penguins into white and black regions based on their spectral characteristics. This segmentation is crucial for subsequent individual identification, enabling specialized processing of different penguin regions. The segmentation results are represented as binary matrices $\mathbf{m}_i \in \mathbb{R}^{h_i \times w_i}$, which we call weighting matrices, where 1 indicates white class pixels and 0 indicates black class pixels. Figure 4 visualizes these weighting matrices $\mathbf{m}_i$, illustrating the separation of white and black regions in the penguin image.

We employ a Gaussian mixture model (GMM)¹⁾ to perform this segmentation by clustering pixel values into two classes, adapting to variations in lighting conditions across images. The GMM is trained to account for these variations in each input cropped region. The classification of each pixel m_{ij} in the weighting matrix is determined as:

$$m_{ij} = \begin{cases} 1 & \text{if } N(x_{ij}; \mu_w, \Sigma_w) > N(x_{ij}; \mu_b, \Sigma_b) \\ 0 & \text{else} \end{cases}, \quad (2)$$

where $j = (u_i, v_i)$ with $u_i \in \{1, \dots, w_i\}$ and $v_i \in \{1, \dots, h_i\}$ denotes the pixel coordinates within the cropped region, representing the width and height dimensions respectively. $N(\cdot)$ represents the Gaussian probability density function, and $\mu_w, \Sigma_w, \mu_b, \Sigma_b$ are the mean and covariance matrices for the white and black classes, respectively.

To determine which GMM class represents the white region, we compare the average pixel values of the two clustered classes. The class with the higher average value is designated as the white class, with its pixels assigned a value of 1 in m_{ij} . Conversely, pixels in the class with the lower average value (representing the black region) are assigned 0 in m_{ij} . This approach ensures that $\mathbf{m}_i$ consistently corresponds to the penguin's white regions, regardless of initial GMM clustering results.

(2) Identification and Ensemble

The Identification and Ensemble component leverages two models for individual penguin identification: one for white and another for black regions. This approach allows us to exploit the distinct spectral characteristics of different parts of a penguin's body.

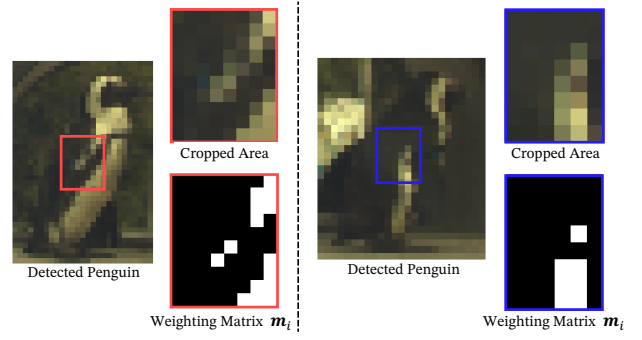


Fig. 4: Visualization of the weighting matrices for a penguin image. White areas in the matrix correspond to pixels with a value of 1, representing the penguin's white regions. Black areas have a value of 0, corresponding to the penguin's black regions.

For each cropped region $\mathbf{X}_i$, we apply two models: $\mathbf{F}^{(w)} : \mathbb{R}^B \rightarrow \mathbb{R}^N$ and $\mathbf{F}^{(b)} : \mathbb{R}^B \rightarrow \mathbb{R}^N$. These models are trained on HS data classified into white and black regions, respectively. Each model takes a B -dimensional HS pixel value as input and outputs an N -dimensional vector, where N corresponds to the number of penguin IDs in the dataset.

To integrate the outputs from these region-specific models, we employ a weighted averaging, using weighting matrices $\mathbf{m}_i$ derived from the Color-Based Pixel Classifier. This method combines the spectral information from both white and black regions based on pixel-wise segmentation. The confidence of the identification for each bounding box is calculated as follows:

$$\mathbf{y}_i = \frac{1}{h_i \times w_i} \sum_{j=1} \{m_{i,j} \mathbf{F}^{(w)}(\mathbf{x}_{i,j}) + (1 - m_{i,j}) \mathbf{F}^{(b)}(\mathbf{x}_{i,j})\}, \quad (3)$$

where $\mathbf{y}_i \in \mathbb{R}^N$ is the confidence vector for the i -th detected penguin, computed as a weighted average of pixel-wise confidence scores within the bounding box.

This ensemble approach ensures that the confidence score for each penguin ID reflects information from both white and black regions, potentially enhancing individual identification accuracy.

(3) Training of Identification Models

The identification models $\mathbf{F}^{(w)}$ and $\mathbf{F}^{(b)}$ are designed to process individual pixels from HS images and predict the corresponding penguin ID. Both models are implemented as multi-layer perceptron (MLP) consisting of 5 fully connected layers. This architecture is adopted from previous work¹⁷⁾, which has shown effectiveness in processing HS data for penguin identification.

To train these models, we utilize the annotated

学習セットに含まれるHS画像から注釈付きバウンディングボックス領域を利用する。学習過程は、我々の提案するシステムと同様の流れに従う:

1) まず、HS画像からCrop Individualsコンポーネントを適用して、関心領域を抽出する。

(2) これらの切り出し領域は、次に次のように処理される。

色ベースのピクセル分類器を用いて、白画素と黒画素を分離する。

3) 分類された画素は、白い領域では D_w の画素、黒い領域では D_b の画素という2つの異なる集合に集められる。

ここで、 $D_w = \{(x^{(w)}_k, t_k)\} \ N_k=1^w$ は白画素の集まりを表し、 $D_b = \{(x^{(b)}_k, t_k)\} \ N_k=1^b$ は黒画素の集まりを表す。インデックス k は、それぞれのセット内のすべてのピクセルにわたって実行される。 D_w に対して、 $x^{(w)}_k \in R^B$ は白に分類されたB次元HSデータ、 t_k は対応するペンギンID、 N_w は白画素の総数である。同様に D_b についても、 $x^{(b)}_k \in R^B$ は黒に分類されたB次元HSデータであり、 N_b は黒画素の総数である。 $F^{(w)}$ と $F^{(b)}$ は、それぞれ白画素と黒画素に対して別々に学習される。このアプローチにより、各モデルはペンギンの白または黒領域の固有のスペクトル特性に特化することができる。

両モデルの学習は、多クラス分類問題として定式化できる。各データセットについて、クロスエントロピー損失 L を最小化する:

$$\mathcal{L} = \sum_{k \in D} CE(t_k, F(x_k)), \quad (4)$$

ここで、 D はどのモデルを学習するかによって D_w か D_b を表し、 x_k は入力画素、 t_k は対応するワンホットエンコードされた真のラベル、 F は $F^{(w)}$ か $F^{(b)}$ のどちらかを表し、 $CE(-, -)$ はクロスエントロピー関数²⁾を表す。この損失関数は、各入力ピクセルを特定のペンギンIDに属するものとして正しく分類するようモデルに促す。

(4) 重複サプレッション

重複抑制コンポーネントは、ペンギン識別システムにおいて重要な後処理ステップとして機能する。その主な目的は、潜在的な衝突を解決し、検出された各ペンギンに固有のIDが割り当てられるようにすることである。このステップは、複数の検出が同じペンギン個体に対応する可能性のある混雑したシーンにおいて特に重要である。

重複抑制のアプローチは、識別とアンサンブルコンポーネントによって生成された信頼度スコアを利用する。

これらのスコアを行列 $Y \in R^{P \times N}$ に整理し、 P は検出されたペンギンの数、 N はデータセット中のペンギンの総数である。 Y の各要素 y_{ik} は、ペンギン検出 i にID k を割り当てるための信頼度スコアを表し、 $i \in \{1, 2, \dots, P\}$ 、 $k \in \{1, 2, \dots, N\}$ である。 i 番目の検出されたペンギンに割り当てられたIDを id_i とする。

重複抑制処理は以下の通りである:

- $L \in R^{P \times N}$ 、ここで l_{ik} はその要素である。

- y_{ik} を降順にソートする。

- y_{ik} を大きいものから小さいものへと繰り返し選択する。

(1) If $l_{ik} = 0$:

- Set id_i to k

- $l_{i*} = 1$ とする(i 行目の全要素)

- Set $l_{*k} = 1$ (all elements in column k)

(2) If $l_{ik} = 1$, skip this entry

- L のすべての要素が 1 に設定されるまで繰り返す。

この二重抑制ステップは、特にペンギンのコロニーが密集しているシナリオにおいて、識別結果の完全性を維持するために極めて重要である。偽陽性を最小化し、我々のシステムが一貫性のある信頼性の高い個々のペンギン識別を提供するのに役立つ。

4. 実験結果

4.1 データセット

実験には、上野動物園²⁴⁾で撮影されたHS画像に基づくカスタムペンギンデータセットを利用した。データセットは、 $N = 16$ 個のペンギンを特徴とする680枚のHS画像から構成される。各HS画像 $H \in R^{H \times W \times B}$ の解像度は 2048×1080 ピクセルで、 $H=2048$ 、 $W=1080$ 、スペクトルバンド $B=151$ は 350nm から 1100nm ⁶⁾の範囲である。

画像は1フレームあたり1~7個のペンギンのグループを捉えている。各画像は、個々のペンギンのバウンディングボックス $b_i = [\text{left } i, \text{top } i, \text{right } i, \text{bottom } i]$ で注釈されており、 $i \in \{1, \dots, P\}$ 、 P は画像中のペンギンの数である。これらのアノテーションから、各ペンギンの切り出しHS画像領域 $X_i \in R^{h_i \times w_i \times B}$ を導出した。ここで、 h_i と w_i はそれぞれ切り出し領域の高さと幅を表す。

解像度の影響を調べるために、バイリニア補間³⁾を用いて、元の解像度の1/4と1/8のHS画像の追加バージョンを用意した。データセットは8:1:1の割合でランダムにトレーニングセット、検証セット、テストセットに分けられた。このデータセットは、ペンギン検出モデル、個人識別モデル F_w 、 F_b を含む、提案システムの学習と評価に使用された。

bounding box regions from the HS images in our training set. The training process follows a similar flow to our proposed system:

- (1) We first apply the Crop Individuals component to extract the regions of interest from the HS images.
- (2) These cropped regions are then processed by the Color Based Pixel Classifier to separate white and black pixels.
- (3) The classified pixels are collected into two distinct sets: pixels of $\mathcal{D}_w$ for white regions and pixels of $\mathcal{D}_b$ for black regions.

Where, $\mathcal{D}_w = \{(\mathbf{x}_k^{(w)}, t_k)\}_{k=1}^{N_w}$ represents the collection of white pixels, and $\mathcal{D}_b = \{(\mathbf{x}_k^{(b)}, t_k)\}_{k=1}^{N_b}$ represents the collection of black pixels. The index k runs over all pixels in each respective set. For $\mathcal{D}_w$, $\mathbf{x}_k^{(w)} \in \mathbb{R}^B$ is a B -dimensional HS data classified as white, t_k is the corresponding penguin ID, and N_w is the total number of white pixels. Similarly for $\mathcal{D}_b$, $\mathbf{x}_k^{(b)} \in \mathbb{R}^B$ is a B -dimensional HS data classified as black, with N_b being the total number of black pixels.

$\mathbf{F}^{(w)}$ and $\mathbf{F}^{(b)}$ are trained separately on these white and black pixels, respectively. This approach allows each model to specialize in the unique spectral characteristics of either the white or black regions of penguins.

The training for both models can be formulated as a multi-class classification problem. We minimize the cross-entropy loss $\mathcal{L}$ for each dataset:

$$\mathcal{L} = \sum_{k \in \mathcal{D}} CE(\mathbf{t}_k, \mathbf{F}(\mathbf{x}_k)), \quad (4)$$

Here $\mathcal{D}$ represents either $\mathcal{D}_w$ or $\mathcal{D}_b$ depending on which model is being trained, $\mathbf{x}_k$ is an input pixel, $\mathbf{t}_k$ is the corresponding one-hot encoded true label, $\mathbf{F}$ represents either $\mathbf{F}^{(w)}$ or $\mathbf{F}^{(b)}$, and $CE(\cdot, \cdot)$ denotes the cross-entropy function²⁾. This loss function encourages the models to correctly classify each input pixel as belonging to a specific penguin ID.

(4) Duplicate Suppression

The duplicate suppression component functions as a critical post-processing step in our penguin identification system. Its primary purpose is to resolve potential conflicts and ensure that each detected penguin is assigned a unique identity. This step is especially crucial in crowded scenes where multiple detections might correspond to the same individual penguin.

Our approach to duplicate suppression utilizes the confidence scores generated by the Identification and

Ensemble component. We organize these scores into a matrix $\mathbf{Y} \in \mathbb{R}^{P \times N}$, where P is the number of detected penguins and N is the total number of penguins in our dataset. Each element y_{ik} of $\mathbf{Y}$ represents the confidence score for assigning ID k to penguin detection i , where $i \in \{1, 2, \dots, P\}$ and $k \in \{1, 2, \dots, N\}$. Let id_i denote the assigned ID for the i -th detected penguin.

The duplicate suppression process is as follows:

- Let $\mathbf{L} \in \mathbb{R}^{P \times N}$, where l_{ik} are its elements.
- Sort y_{ik} in descending order.
- Iteratively select y_{ik} from largest to smallest:
 - (1) If $l_{ik} = 0$:
 - Set id_i to k
 - Set $l_{i*} = 1$ (all elements in row i)
 - Set $l_{*k} = 1$ (all elements in column k)
 - (2) If $l_{ik} = 1$, skip this entry
- Repeat until all elements of $\mathbf{L}$ are set to 1

This duplicate suppression step is crucial for maintaining the integrity of our identification results, especially in scenarios with densely populated penguin colonies. It helps to minimize false positives and ensures that our system provides consistent and reliable individual penguin identifications.

4. Experiments

4.1 Dataset

For our experiments, we utilized a custom penguin dataset based on HS images captured at Ueno Zoo²⁴⁾. The dataset consists of 680 HS images featuring $N = 16$ individual penguins. Each HS image $\mathbf{H} \in \mathbb{R}^{H \times W \times B}$ has a resolution of 2048×1080 pixels with $H = 2048$, $W = 1080$, and 151 spectral bands $B = 151$ ranging from 350 nm to 1100 nm⁶⁾.

The images capture groups of 1 to 7 penguins per frame. Each image is annotated with bounding boxes $\mathbf{b}_i = [\text{left}_i, \text{top}_i, \text{right}_i, \text{bottom}_i]$ for individual penguins, where $i \in \{1, \dots, P\}$ and P is the number of penguins in the image. From these annotations, we derived cropped HS image regions $\mathbf{X}_i \in \mathbb{R}^{h_i \times w_i \times B}$ for each penguin, where h_i and w_i represent the height and width of each cropped region, respectively.

To investigate resolution effects, we prepared additional versions of the HS images at 1/4 and 1/8 of the original resolution using bilinear interpolation³⁾. The dataset was randomly divided into training, validation, and test sets in an 8:1:1 ratio. This dataset was used for training and evaluating our proposed system, including the penguin detection model and individual identification models $\mathbf{F}^w$ and $\mathbf{F}^b$.

4.2 実装の詳細

実験は、Intel Core i9-13900K CPU、128GB RAM、NVIDIA GeForce RTX A6000 GPUを搭載したシステムで、PyTorch¹⁹⁾を使用して実施した。RGBに変換されたHS画像に対して、軽量なYOLOモデル⁸⁾を微調整した。学習はSGD¹²⁾を利用し、学習率は 1×10^{-2} 、100エポック、バッチサイズは16である。オリジナル、1/4、1/8解像度で別々のモデルを学習した。

我々のピクセル単位の分類モデルは、ReLU活性化とバッチ正規化を持つ5層MLPを採用した。ドロップアウトは、最終層の確率を0.5、最終層の確率を0.1として実施した。このモデルは、Adam optimizer⁹⁾を用いて、ピクセル集合 D_w と D_b で 300 エポック学習した。初期学習率は 1×10^{-3} とし、全エポックの 1/10 ごとに 0.6 ずつ減少させた。学習にはバッチサイズ128を使用した。YOLOモデルに採用されたアプローチに従い、全てのスケールで性能を最大化するために、ピクセル単位の分類モデルの解像度固有のインスタンスを開発し、訓練した。

ノイズを低減するため、画素抽出前のHS画像に平滑化フィルタを適用した：全解像度画像には 5×5 フィルタ、1/4解像度画像と1/8解像度画像には 3×3 フィルタ。すべての解像度において、ボックスの寸法の $r = 0.3$ で定義される注釈付きバウンディングボックスの中心領域からピクセルを選択した。この r の値は、十分なペンギン固有のピクセルを含めることと、背景の干渉を最小化することのバランスをとるために経験的に決定された。システムの性能は r の小さな変動に対して相対的なロバスト性を示したが、適切な値を維持することは、特に効果的なアンサンブル処理を保証する上で、最適な結果を得るために重要である。

本システムの性能を評価するために、平均平均精度(mAP)、精度、再現率、F1スコアという複数の指標を採用した。各ペンギンIDについて、Intersection over Union (IoU)が50[%]以上の検出を真陽性とし、平均精度(AP)を算出した。mAPは、すべてのペンギンIDのAP値を平均し、ローカライゼーション精度と同定の一貫性の両方を評価することで計算された。

Precision、Recall、F1-scoreの計算では、信頼閾値を0.45とした。Precisionは全陽性予測のうち正しい陽性予測の割合を測定し、Recallは実際に正しく識別された陽性ケースの割合を表す。

表1:様々な解像度のHS画像に対する、微調整されたYOLOモデルを用いたペンギン検出の性能評価。表は、様々なIntersections over Union (IoU)の閾値における平均精度(AP)を示している。

IOU threshold	Resolution		
	Full	1/4	1/8
IoU ₅₀	96.90	97.93	97.85
IoU ₇₅	96.68	97.74	97.47
IoU ₉₀	95.99	97.04	96.08
IoU ₉₅	88.38	88.98	86.02

PrecisionとRecallは、モデルの精度のバランスの取れた尺度を提供する。

4.3 前提条件

様々な距離での検出をシミュレートするために、完全解像度、1/4解像度、1/8解像度の画像でペンギン検出モデルを評価した。表1は、これらの解像度におけるYOLOモデルの異なるIoU閾値でのAPスコアを示している。IoUの閾値を変化させ、緩い物体定位基準から厳しい物体定位基準まで、モデルの性能を評価した。

このモデルは、様々なIoU閾値と画像解像度において卓越した性能を示している。IoU₅₀では、96.90[%]、97.93[%]、97.85[%]と、それぞれフル解像度、1/4解像度、1/8解像度で著しく高いAPスコアを達成している。このロバストな性能は、より厳しい閾値でも持続し、それぞれの解像度のIoU₉₀でのAPスコアは95.99[%]、97.04[%]、96.08[%]、IoU₉₅でのAPスコアは88.38[%]、88.98[%]、86.02[%]であった。

これらの結果は、様々な画像解像度のペンギン検出において、我々の微調整したYOLOモデルの有効性を強調するものである。細かいディテールが減少しても高い性能を維持するモデルの能力は、RGBの外観情報が複雑な群れシナリオにおける頑健なペンギン検出のための十分な手がかりを提供することを示唆している。様々な解像度と厳密なIoU閾値で一貫した性能を発揮することは、モデルの回復力と多様な実世界条件での適用可能性を示している。

4.4 アブレーション研究

我々の貢献を検証するために、2つの重要な革新に焦点を当てたアブレーション研究を行った：識別とアンサンブル、重複抑制である。表2は、全解像度画像と1/8解像度画像の両方に対する結果を示している。

本システムは、ペンギンの白と黒の領域に対して特別に訓練された2つのモデルを採用している。Identification and Ensembleアプローチは、

4.2 Implementation Details

Experiments were conducted on a system with an Intel Core i9-13900K CPU, 128 GB RAM, and an NVIDIA GeForce RTX A6000 GPU, using PyTorch¹⁹⁾.

We fine-tuned the lightweight YOLO model⁸⁾ on our HS images converted to RGB. The training utilized SGD¹²⁾ with a learning rate of 1×10^{-2} for 100 epochs and a batch size of 16. Separate models were trained for original, 1/4, and 1/8 resolutions.

Our pixel-wise classification model employed a 5-layer MLP with ReLU activations and batch normalization. Dropout was implemented using probabilities of 0.5 for the final layer and 0.1 for the penultimate layer. This model was trained on pixel sets D_w and D_b for 300 epochs using Adam optimizer⁹⁾. The initial learning rate was set to 1×10^{-3} , decaying by a factor of 0.6 every 1/10 of total epochs. We used a batch size of 128 for training. Following the approach employed for the YOLO models, we developed and trained resolution-specific instances of the pixel-wise classification model to maximize performance across all scales.

To reduce noise, smoothing filters were applied to the HS images before pixel extraction: a 5×5 filter for full-resolution images, and 3×3 filters for 1/4 and 1/8-resolution images. Across all resolutions, we selected pixels from a central region of the annotated bounding boxes, defined by $r = 0.3$ of the box dimensions. This value of r was empirically determined to provide a good balance between including sufficient penguin-specific pixels and minimizing background interference. While the system's performance showed relative robustness to small variations in r , maintaining an appropriate value is important for optimal results, particularly in ensuring effective ensemble processing.

To assess the performance of our system, we employed multiple metrics: mean Average Precision (mAP), Precision, Recall, and F1-score. We calculated Average Precision (AP) for each penguin ID, considering detections with an Intersection over Union (IoU) of at least 50[%] with the ground truth bounding boxes as true positives. The mAP was computed by averaging AP values across all penguin IDs, assessing both localization accuracy and identification consistency.

For Precision, Recall, and F1-score calculations, we adopted a confidence threshold of 0.45. Precision measures the proportion of correct positive predictions among all positive predictions, while Recall represents the proportion of actual positive cases that were correctly identified. The F1-score, the harmonic mean of

Table 1: Performance evaluation of penguin detection using fine-tuned YOLO models on HS images of various resolutions. The table shows Average Precision (AP) at various Intersections over Union (IoU) thresholds.

IOU threshold	Resolution		
	Full	1/4	1/8
IoU ₅₀	96.90	97.93	97.85
IoU ₇₅	96.68	97.74	97.47
IoU ₉₀	95.99	97.04	96.08
IoU ₉₅	88.38	88.98	86.02

Precision and Recall, provides a balanced measure of the model's accuracy.

4.3 Preliminaries

We evaluated our penguin detection model across full, 1/4, and 1/8-resolution images to simulate detection at various distances. Table 1 presents the AP scores at different IoU thresholds for the YOLO model across these resolutions. We varied the IoU thresholds to assess the model's performance from loose to strict object localization criteria.

The model demonstrates exceptional performance across various IoU thresholds and image resolutions. At IoU₅₀, it achieves remarkably high AP scores: 96.90[%], 97.93[%], and 97.85[%] for full, 1/4, and 1/8 resolutions, respectively. This robust performance persists even at more stringent thresholds, with AP scores at IoU₉₀ of 95.99[%], 97.04[%], and 96.08[%], and at IoU₉₅ of 88.38[%], 88.98[%], and 86.02[%] for the respective resolutions.

These results underscore the efficacy of our fine-tuned YOLO model in penguin detection across a spectrum of image resolutions. The model's ability to maintain high performance even when fine details are reduced suggests that RGB appearance information provides sufficient cues for robust penguin detection in complex flock scenarios. This consistent performance across varying resolutions and strict IoU thresholds demonstrates the model's resilience and applicability in diverse real-world conditions.

4.4 Ablation Study

To verify our contributions, we conducted an ablation study focusing on two key innovations: Identification and Ensemble, and duplicate suppression. Table 2 presents the results for both full-resolution and 1/8-resolution images.

Our system employs two models specifically trained for the white and black regions of penguins. The Identification and Ensemble approach incorporates re-

表2: 重複抑制の有無による識別方法の違いを比較したアブレーション研究の結果。性能は、フル解像度画像と1/8解像度画像のmAP、Precision、Recall、F1スコアを用いて測定。

同定方法について	Duplicate Suppression	Full-resolution				1/8-resolution			
		mAP	Precision	Recall	F1-score	mAP	Precision	Recall	F1-score
White and Black Regions	×	68.28	0.67	0.78	0.72	66.35	0.65	0.76	0.70
White Region	×	66.66	0.70	0.77	0.73	69.19	0.66	0.78	0.71
Black Region	×	70.92	0.69	0.79	0.74	71.47	0.68	0.78	0.72
識別とアンサンブル	×	72.11	0.70	0.80	0.75	74.19	0.70	0.82	0.75
白人と黒人の地域	✓	69.27	0.67	0.79	0.72	68.24	0.67	0.78	0.71
White Region	✓	68.00	0.70	0.79	0.74	70.11	0.67	0.78	0.72
黒人地域	✓	72.25	0.70	0.80	0.75	73.57	0.71	0.80	0.75
識別とアンサンブル	✓	72.70	0.70	0.81	0.75	75.94	0.71	0.83	0.76

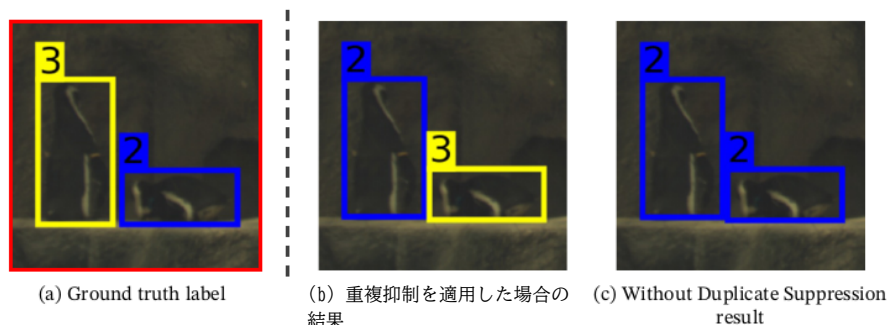


図5: 全解像度のテストサンプルにおけるペンギン識別結果。(a) グラウンドトゥールスラベル。(b) 重複抑制を適用した場合。(c) 重複抑制なし。この例は、重複抑制がいかに誤った識別につながるかを示している。ID2の信頼度スコアは、左ペンギンが0.56、右ペンギンが0.52である。

両方の領域からの結果を取り入れ、「White Region」と「Black Region」のバリエーションは、それぞれの領域からの結果を使用する。白と黒の領域」法は、両方の領域で同時に学習した単一のモデルを使用する。

全解像度画像において、Identification and Ensembleアプローチは、重複抑制ありで72.70[%]のmAPと最高の精度を達成し、「White and Black Regions」アプローチ(69.27[%]mAP)を3.43[%]上回った。この改善は、ペンギン識別のために白と黒の領域を独立して処理することの有効性を強調している。

白色領域のみを使用した場合は68.00[%]mAP、黒色領域のみを使用した場合は72.25[%]mAPとなった。このことは、黒い領域が、個人識別のためのより識別的な特徴を含んでいる可能性を示唆している。複合的なアプローチは72.70[%]のmAPを達成し、単一領域の手法を上回り、白い領域で補完的な情報を示している。

重複抑制はすべての手法で有効であり、mAPでは0.59[%]から1.34[%]の範囲で改善した。注目すべきは、「White Region」バリエーションは重複抑制で1.34[%]の改善を示し、66.66[%]から68.00[%]のmAPを示したことである。

全解像度画像で観察された性能傾向は、1/8解像度画像でも悪化しなかった。

Identification and Ensembleアプローチは引き続き他の手法を凌駕し、75.94[%]のmAPを達成し、重複抑制を実現した。このことは、本手法が遠方の被写体を撮影することによる画質劣化に対してロバストであることを示唆している。

しかし、重複抑制が時折誤った予測につながる可能性があることに注意することが重要である。図5に示すように、重複抑制処理により誤認識が生じることがある。重複抑制(c)がない場合、一方のペンギンは正しく識別され、他方のペンギンは同じIDを持つと誤って予測される。重複抑制を適用した場合(b)、両者に異なるペンギンIDを誤って割り当ててしまい、どちらもグラウンドトゥールスラベルと一致しない(a)。左ペンギンと右ペンギンの信頼度は、それぞれID 2で0.56と0.52である。重複抑制により、ID 2は左ペンギンの方が若干スコアが高く、ID 3は右ペンギンに割り当てられる。しかし、グラウンドトゥールスでは、右のペンギンが正しいID 2であることがわかる。

図5は、この方法が失敗したケースを示しているが、これらの結果は、識別とアンサンブルと重複抑制の両方が、我々のペンギン識別システムの性能を向

Table 2: Ablation study results comparing different identification methods with and without duplicate suppression. Performance is measured using mAP, Precision, Recall, and F1-score for full and 1/8-resolution images.

Identification method	Duplicate Suppression	Full-resolution				1/8-resolution			
		mAP	Precision	Recall	F1-score	mAP	Precision	Recall	F1-score
White and Black Regions	×	68.28	0.67	0.78	0.72	66.35	0.65	0.76	0.70
White Region	×	66.66	0.70	0.77	0.73	69.19	0.66	0.78	0.71
Black Region	×	70.92	0.69	0.79	0.74	71.47	0.68	0.78	0.72
Identification and Ensemble	×	72.11	0.70	0.80	0.75	74.19	0.70	0.82	0.75
White and Black Regions	✓	69.27	0.67	0.79	0.72	68.24	0.67	0.78	0.71
White Region	✓	68.00	0.70	0.79	0.74	70.11	0.67	0.78	0.72
Black Region	✓	72.25	0.70	0.80	0.75	73.57	0.71	0.80	0.75
Identification and Ensemble	✓	72.70	0.70	0.81	0.75	75.94	0.71	0.83	0.76

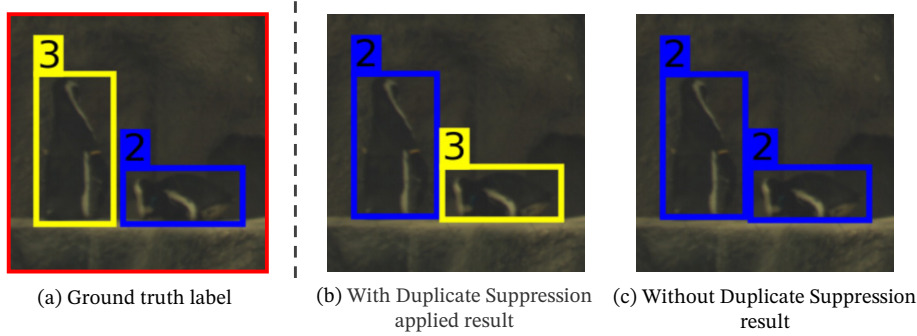


Fig. 5: Penguin identification results on a full-resolution test sample. (a) Ground truth label. (b) With duplicate suppression applied. (c) Without duplicate suppression. This example demonstrates how duplicate suppression can lead to incorrect identification. The confidence scores for ID 2 are 0.56 for the left penguin and 0.52 for the right penguin.

sults from both regions, while the "White Region" and "Black Region" variants use results from the respective regions. The "white and black Regions" method uses a single model trained on both regions simultaneously.

For full-resolution images, the Identification and Ensemble approach achieved the highest accuracy with 72.70[%] mAP with duplicate suppression, surpassing the "White and Black Regions" approach (69.27[%] mAP) by 3.43[%]. This improvement underscores the effectiveness of independently processing the white and black regions for penguin identification.

Using only the white region yielded 68.00[%] mAP, while the black region alone produced 72.25[%] mAP. This suggests the black region may contain more discriminative features for individual identification. The combined approach achieved 72.70[%] mAP, outperforming both single-region methods and indicating complementary information in the white region.

Duplicate suppression proved effective across all methods, with improvements ranging from 0.59[%] to 1.34[%] in mAP. Notably, the "White Region" variant showed a 1.34[%] improvement with duplicate suppression, from 66.66[%] to 68.00[%] mAP.

The performance trends observed in full-resolution

images did not deteriorate in 1/8-resolution images. The Identification and Ensemble approach continued to outperform other methods, achieving 75.94[%] mAP with duplicate suppression. This suggests that our method is robust to image quality degradation from photographing distant subjects.

However, it is important to note that duplicate suppression can occasionally lead to incorrect predictions. As illustrated in Figure 5, the duplicate suppression process can sometimes result in misidentification. In the example shown, without duplicate suppression (c), one penguin is correctly identified while the other is incorrectly predicted as having the same ID. When duplicate suppression is applied (b), it incorrectly assigns different penguin IDs to both individuals, neither of which matches the ground truth label (a). The left and right penguins have confidence scores of 0.56 and 0.52 respectively for ID 2. Duplicate suppression assigns ID 2 to the left penguin due to its slightly higher score and ID 3 to the right penguin. However, the ground truth shows the right penguin is the correct ID 2.

Although Figure 5 shows a case where the method fails, these findings indicate that both Identification and Ensemble and duplicate suppression improve our pen-

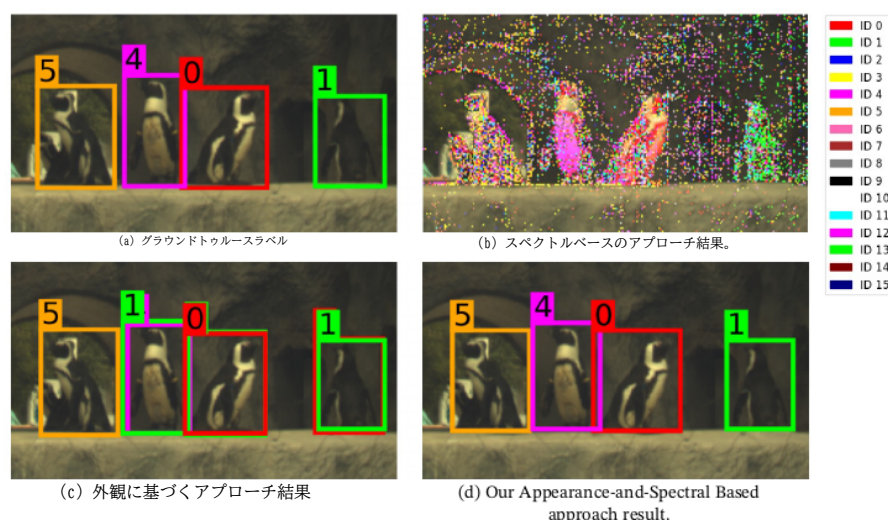


図6:異なるアプローチによるペンギン識別結果の視覚的比較。本手法は、様々な画像解像度において、個々のペンギンを正確に識別する上で優れた性能を発揮する。Spectralbasedアプローチでは、色分けされた領域は予測されたペンギンの位置を示す:正しい識別を緑、誤識別を赤で示す。

上させることを示している。

4.5 我々のシステムの評価

提案システムを評価するために、既存のシングルモダリティアプローチと本手法を比較した。我々の知る限り、ペンギンの個体識別のためにスペクトル情報と外観情報を組み合わせた既存のシステムは存在しない。我々の分析は、主に外観ベースとスペクトルベースのアプローチを考慮し、また、遠くの群れの観測をシミュレートするために、より低解像度のHS画像を用いた性能も考慮する。

外観に基づくアプローチは、高速で広く使われている物体検出モデルであるYOLOを採用している。YOLOモデルは、ペンギンのローカライズと個人のアイデンティティの割り当てという2つのタスクを実行するように訓練されている。これに対して、我々の手法は、ローカライゼーションにのみYOLOを用いる。スペクトルベースのアプローチは、2段階のMLPプロセスを採用している。第一段階では、ピクセルをペンギンか背景のどちらかに分類するバイナリセグメンテーションを行う。次に第2段階では、セグメント化されたペンギンピクセルの個々のIDを予測する。本手法では、画素レベルのID予測に単一のMLPを用いる。

これらの分析は、群れの画像からペンギンの個体をすべて特定するという、我々の新しいタスクのベンチマークとなる。既存のアプローチを修正した比較評価を通じて、ペンギン識別におけるこの未解明の課題に対処するための我々のシステムの有効性を実証することを目的とする。

表3:異なるアプローチを用いたペンギン個体識別のmAP、Precision、Recall、F1-scoreの画像解像度間の比較。

Method	Resolution	mAP	Precision	Recall	F1-score
アビאלランス ベース	Full	36.12	0.57	0.24	0.34
	1/4	41.46	0.58	0.33	0.42
	1/8	36.40	0.53	0.28	0.37
Spectral based	Full	5.88	0.05	0.05	0.05
	1/4	6.13	0.07	0.53	0.08
	1/8	5.95	0.06	0.23	0.06
外観とスペク トルに基づく (我々の)	Full	72.70	0.70	0.81	0.74
	1/4	76.82	0.72	0.83	0.77
	1/8	75.94	0.71	0.83	0.76

4.6 評価結果

表3は、様々な画像解像度における、我々の提案システムと既存のアプローチとの比較分析である。我々のシステムは、すべての解像度で70.00[%]を超えるmAPスコアを一貫して達成し、ペンギンの個体識別と位置特定において高い精度と信頼性を示した。このパフォーマンスレベルは、野生生物のモニタリングや保全活動において、このシステムを効果的に利用できる可能性を示しており、研究者はペンギンの個体群を最小限の手作業で追跡することができます¹⁵⁾。

一方、従来の外観に基づくアプローチでは、mAPスコアが36.12[%]から41.46[%]と一貫性のない性能を示した。スペクトルベースのアプローチは、mAPスコアが6[%]前後と、さらに限定的な有効性を示した。これらの結果は、群れ内の個々のペンギンを識別する上で、我々の提案するシステムが頑健であることと、

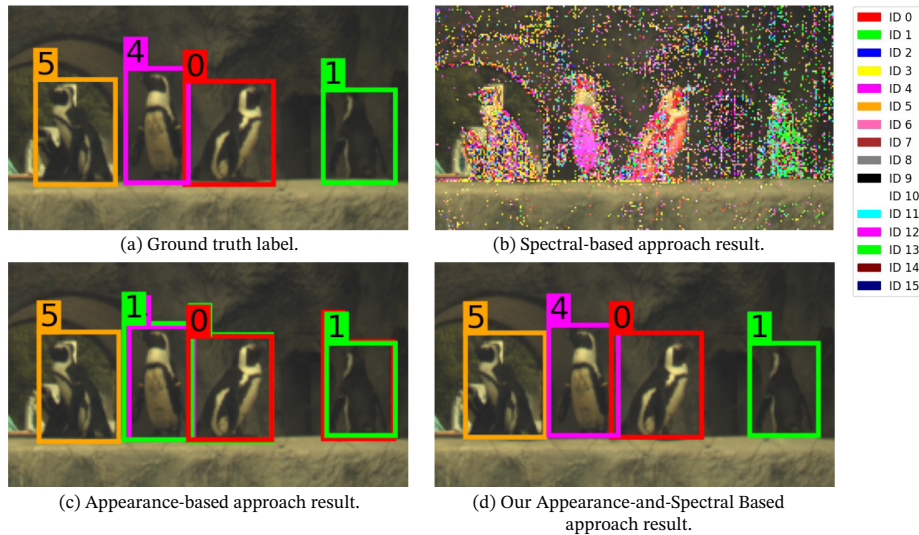


Fig. 6: Visual comparison of penguin identification results across different approaches. Our method demonstrates superior performance in accurately identifying individual penguins across various image resolutions. For the Spectral-based approach, colored areas indicate predicted penguin locations: green for correct identifications and red for misidentifications.

guin identification system’s performance.

4.5 Evaluation of Our System

We compared our method with existing single-modality approaches as benchmarks to evaluate our proposed system. To the best of our knowledge, no existing system combines spectral and appearance information for individual identification of penguins. Our analysis primarily considers appearance-based and spectral-based approaches, as well as a performance with lower-resolution HS images to simulate distant flock observations.

The appearance-based approach employs YOLO, a fast and widely used object detection model. The YOLO model is trained to perform dual tasks: localizing penguins and assigning individual identities. Our method, in contrast, uses YOLO only for localization. The spectral-based approach employs a two-stage MLP process. The first stage performs binary segmentation, classifying pixels as either penguin or background. The second stage then predicts individual IDs for the segmented penguin pixels. Our method uses a single MLP for pixel-level ID prediction.

These analyses benchmark our novel task of identifying all individual penguins in a flock image. Through a comparative evaluation with modified existing approaches, we aim to demonstrate our system’s effectiveness in addressing this unexplored challenge in penguin identification.

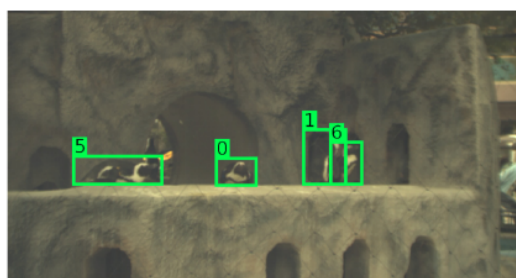
Table 3: Comparison of mAP, Precision, Recall, and F1-score for penguin individual identification using different approaches across image resolutions.

Method	Resolution	mAP	Precision	Recall	F1-score
Appearance based	Full	36.12	0.57	0.24	0.34
	1/4	41.46	0.58	0.33	0.42
	1/8	36.40	0.53	0.28	0.37
Spectral based	Full	5.88	0.05	0.05	0.05
	1/4	6.13	0.07	0.53	0.08
	1/8	5.95	0.06	0.23	0.06
Appearance and-Spectral based(Ours)	Full	72.70	0.70	0.81	0.74
	1/4	76.82	0.72	0.83	0.77
	1/8	75.94	0.71	0.83	0.76

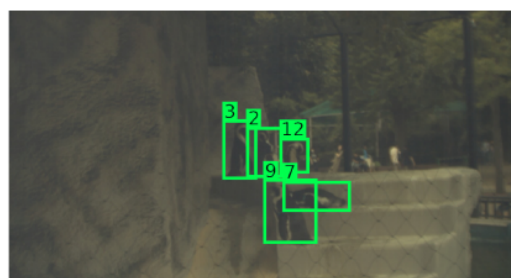
4.6 Evaluation Results

Table 3 presents a comparative analysis of our proposed system versus existing approaches at various image resolutions. Our system consistently achieved mAP scores that exceeded 70.00[%] for all resolutions, demonstrating high precision and reliability in identifying and localizing individual penguins. This performance level indicates the potential for effective use of the system in wildlife monitoring and conservation efforts, allowing researchers to track penguin populations with minimal manual intervention¹⁵.

In contrast, the conventional appearance-based approach showed inconsistent performance, with mAP scores ranging from 36.12[%] to 41.46[%]. The spectral-based approach demonstrated even more limited efficacy, with mAP scores around 6[%]. These results underscore the robustness of our proposed system in identifying individual penguins within flocks and their re-



(a) サンプル群れ画像1



(b) サンプル群れ画像2



(c) サンプル群れ画像3



(d) サンプル群れ画像4

図7:様々な画像におけるペンギンの個体識別の結果。緑色のバウンディングボックスは正しい識別、赤色のボックスは誤識別、紫色のボックスは検出エラーを示す。本システムは、多様な姿勢や部分的な視点からペンギンを識別する際の頑健性を示している。注目すべき困難なケースとして、(d)のペンギンID 7があり、部分的なビューしか利用できない。(c)ではID11とID0の誤認識があり、誤認識が見られる。

解像度の劣化に対する耐性を強調するものである。

図6は、異なるアプローチによる出力を視覚的に比較したものである。RGBアプローチは、比較的高いmAPにもかかわらず、1つのペンギンに複数のバウンディングボックスを割り当てることが多く、同じIDで異なるペンギンをラベル付けすることが多い。このことは、mAPスコアが、個体を正確に特定することなく、画像に存在するペンギンIDを予測するモデルの能力を反映している可能性を示唆している。

HSアプローチは、ペンギンと背景を正確に区別することに限界がある。しかし、適切に検出されたペンギン地域内の個体を正しく識別する上で有望な結果を示し、正確な位置情報が提供された場合の個体識別の可能性を示している。

提案システムは、RGBアプローチとHSアプローチの相補的な長所を活用することで、RGBアプローチとHSアプローチの両方の限界を克服し、すべての個々のペンギンを識別することに成功した。

提案システムの性能を可視化するために、図7にシステムの検出と識別の結果を反映したバウンディングボックスを持つ出力画像のサンプルを示す。各バウンディングボックスは、左上隅に予測されたペンギンIDを表示する。

緑、赤、紫のバウンディングボックスは、それぞれ正しい識別、正しくない識別、検出されなかったペンギンを示す。これらの画像は、姿勢の変化や部分的な視界の変化といった困難なシナリオにおいても、群れの中で複数のペンギンを識別するシステムの有効性を示している。例えば、図7(d)では、ペンギンID 7は、上半身のみが段差の上に見えるにもかかわらず、正しく識別されており、地形の特徴に起因する部分的なビューを管理するシステムの能力を示している。

5. おわりに

本論文では、外観情報とスペクトル情報を組み合わせた群れ画像におけるペンギン個体識別のための新しいシステムを紹介する。我々のアプローチは、ペンギン検出のためのRGB画像からの外観情報と、個人識別のためのHSデータからのスペクトル情報を活用する。主な革新点は、白と黒の領域を別々に処理するアンサンブル法と、重複抑制技術である。

実験により、本システムが外観のみ、スペクトルのみのアプローチよりも優れていることが実証され、異なる解像度で平均平均70.00%以上の精度を達成した。本システムは、様々なペンギンの姿勢や部分的な視界に対して堅牢性を示した。

今後の研究では、照明、天候、

【機械翻訳コンテンツの著作権について】

当サイトに掲載されている原著論文の著作権は映像情報メディア学会に帰属します。原著論文および翻訳論文の無断使用は禁止します。
引用の際には、必ず原著論文の書誌情報をご記載ください。詳細は映像情報メディア学会著作権規定をご覧ください。

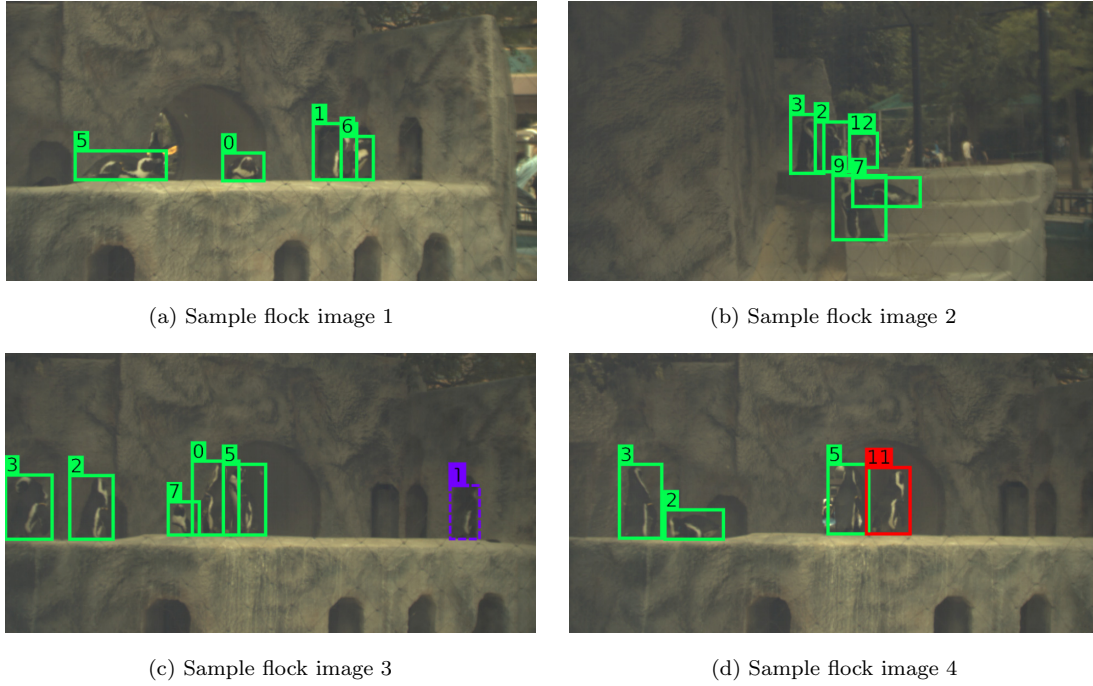


Fig. 7: Results of penguin individual identification across various images. Green bounding boxes indicate correct identifications, red boxes show misidentifications and purple boxes represent detection errors. The system demonstrates robustness in identifying penguins in diverse postures and partial views. Notable challenging cases include penguin ID 7 in (d), where only partial views are available. Incorrect identification is noted in (c) with a misidentification between ID11 and ID0.

silience to resolution degradation.

Figure 6 provides a visual comparison of outputs from different approaches. The RGB approach, despite its relatively higher mAP, often assigns multiple bounding boxes to a single penguin and frequently labels different penguins with the same ID. This suggests that its mAP score may reflect the model’s ability to predict penguin IDs present in the image without accurately localizing individuals.

The HS approach shows limitations in accurately distinguishing penguins from the background. However, it demonstrates promising results in correctly identifying individuals within properly detected penguin regions, indicating its potential for individual identification when accurate location information is provided.

Our proposed system successfully identified all individual penguins, overcoming the limitations of both RGB and HS approaches by leveraging their complementary strengths.

To visualize the performance of our proposed system, Figure 7 presents sample output images with bounding boxes reflecting the system’s detection and identification results. Each bounding box displays the predicted penguin ID in its upper left corner. Green, red, and purple bounding boxes indicate correct identifications,

incorrect identifications, and undetected penguins, respectively. These images demonstrate the system’s effectiveness in identifying multiple penguins within flocks, even in challenging scenarios such as varying postures or partial views. For example, in Figure 7(d), penguin ID 7 is correctly identified despite only its upper body being visible above a step, showcasing the system’s capability to manage partial views caused by terrain features.

5. Conclusion

This paper presented a novel system for individual penguin identification in flock images combining appearance and spectral information. Our approach leverages appearance information from RGB images for penguin detection and spectral information from HS data for individual identification. Key innovations include an ensemble method that processes white and black regions separately and a duplicate suppression technique.

Experiments demonstrated our system’s superiority over appearance-only and spectral-only approaches, achieving over 70.00% mean Average Precision across different resolutions. The system showed robustness in various penguin postures and partial views.

In our future work, we will test the system’s perfor-

カメラのセットアップが異なる多様な環境において、システムの性能をテストする予定である。これは、その実用性を評価し、改善すべき点を特定するのに役立つ。

私たちは、私たちのシステムの設計が、キリンやバライス・クロノース・ヒツジなど、さまざまな種に適応できる可能性があると考えている。異なる動物種について他のデータセットを構築し、我々のシステムを他の種に適用する。この研究は、多様な生態学的状況において、カメラを用いた野生生物モニタリング技術の普及につながるものである。

Appendix

We thank and honor the Ueno Zoological Gardens²⁴⁾ for allowing us to capture HS images. This work was supported by JSPS KAKENHI 22K12081.

References

- 1) Bishop, C. M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 1 edition (2007).
- 2) Bishop, C. M. and Bishop, H.: *Deep learning: Foundations and concepts*, Springer Nature, Cham, Switzerland (2023).
- 3) Bradski, G.: The OpenCV Library (2000).
- 4) Burghardt, T., Thomas, B., Barham, P. J. and Calic, J.: Automated visual recognition of individual African penguins, *Fifth International Penguin Conference* (2004).
- 5) Dodd Jr, C. K.: Infrared reflectance in chameleons (Chamaeleonidae) from Kenya (1981).
- 6) Eba Japan Co.Ltd.: Hyper spectral camera NH9, <https://ebajapan.jp/products/hyper-spectral-camera/>.
- 7) Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M. and Grieve, B.: Towards on-farm pig face recognition using convolutional neural networks, *Computers in Industry*, Vol. 98, pp. 145–152 (2018).
- 8) Jocher, G.: Ultralytics YOLOv5 (2020).
- 9) Kingma, D. P.: Adam: A method for stochastic optimization (2014).
- 10) Kolmann, M., Kalacska, M., Lucanus, O., Sousa, L., Wainwright, D., Arroyo-Mora, J. and Andrade, M.: Hyperspectral data as a biodiversity screening tool can differentiate among diverse Neotropical fishes, *Scientific reports*, Vol. 11, No. 1, p. 16157 (2021).
- 11) Kumar, S. and Singh, S. K.: Visual animal biometrics: survey, *Iet Biometrics*, Vol. 6, No. 3, pp. 139–156 (2017).
- 12) Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324 (1998).
- 13) Magnusson, M., Sigurdsson, J., Armansson, S. E., Ulfarsson, M. O., Deborah, H. and Sveinsson, J. R.: Creating RGB Images from Hyperspectral Images Using a Color Matching Function, *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2045–2048 (2020).
- 14) Matkowsky, W. M., Kong, A. W. K., Su, H., Chen, P., Hou, R. and Zhang, Z.: Giant Panda Face Recognition Using Small Dataset, *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE (2019).
- 15) Matkowsky, W. M., Kong, A. W. K., Su, H., Chen, P., Hou, R. and Zhang, Z.: Giant panda face recognition using small dataset, *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 1680–1684 (2019).
- 16) Mielewicz, M., Liebis, F., Walter, A. and Greven, H.: Near-infrared (NIR)-reflectance in insects-Phenetic studies of 181 species, *Entomologie heute*, Vol. 24, pp. 183–215 (2012).
- 17) Noboru, Y., Ozasa, Y. and Tanaka, M.: Hyperspectral Image Dataset for Individual Penguin Identification, *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 9383–9387 (2024).
- 18) Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jovic, N. and Clune, J.: A deep active learning system for species identification and counting in camera trap images, *Methods in ecology and evolution*, Vol. 12, No. 1, pp. 150–161 (2021).

- 19) Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., pp. 8024–8035 (2019).
- 20) Pinto, F., Mielewicz, M., Liebis, F., Walter, A., Greven, H. and Rascher, U.: Non-invasive measurement of frog skin reflectivity in high spatial resolution using a dual hyperspectral approach, *PLoS One*, Vol. 8, No. 9, p. e73234 (2013).
- 21) Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D. and Carvalho, S.: Chimpanzee face recognition from videos in the wild using deep learning, *Science Advances*, Vol. 5, No. 9, p. eaaw0736 (2019).
- 22) Schwalm, P. A., Starrett, P. H. and McDiarmid, R. W.: Infrared reflectance in leaf-sitting neotropical frogs, *Science*, Vol. 196, No. 4295, pp. 1225–1226 (1977).
- 23) Sherley, R., Burghardt, T., Barham, P., Cuthill, I. and Campbell, N.: Spotting the difference: towards fully-automated population monitoring of African penguins *Spheniscus demersus*, *Endangered Species Research*, Vol. 11, No. 2, pp. 101 – 111 (2010).
- 24) Tokyo Zoological Park Society: Ueno Zoological Gardens, <https://www.tokyo-zoo.net/english/ueno/index.html>.
- 25) Wu, J., Xu, W., He, J. and Lan, M.: YOLO for Penguin Detection and Counting Based on Remote Sensing Images, *Remote Sensing*, Vol. 15, No. 10, p. 2598 (2023).



Youta Noboru Youta Noboru received the B.Eng. degree in Information System Engineering from Tokyo Denki University in 2024. He is currently pursuing the Master degree in Machine Learning and Computer Vision at Tokyo Denki University. His research interests include visual perception, and hyperspectral sensing.



Yuko Ozasa Yuko Ozasa received her PhD degree in engineering from Kobe University in 2015. She was a postdoctoral researcher at National Institute of Advanced Industrial Science and Technology (AIST) in 2015. She had been a research associate at Graduate School of Science and Technology, Keio University from 2016 to 2020. She had been an assistant professor at Graduate School of System Design and Technology, Tokyo Denki University from 2020 to 2022. Since 2023, she has been an associate professor at the university. She has also been a guest professor at Gakushuin University since 2024. Her research interests include object recognition and grounding, visual perception, and hyperspectral sensing.



Masayuki Tanaka Masayuki Tanaka received his bachelor's and master's degrees in control engineering and a doctoral degree from Tokyo Institute of Technology in 1998, 2000, and 2003. He was a software engineer at Agilent Technologies from 2003 to 2004. He was a Research Scientist at Tokyo Institute of Technology from 2004 to 2008. He was an Associate Professor at the Graduate School of Science and Engineering, Tokyo Institute of Technology, from 2008 to 2016. He was a Visiting Scholar at Department of Psychology, Stanford University, from 2013 to 2014. He was an Associate Professor at School of Engineering, Tokyo Institute of Technology, from 2016 to 2017. He was a Senior Researcher at National Institute of Advanced Industrial Science and Technology from 2017 to 2020. He was an Associate Professor at School of Engineering, Tokyo Institute of Technology, from 2020 to 2023. Since 2023, he has been a Professor at Graduate Major in Engineering Sciences and Design, Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology. In 2024, following the university's integration, he became a Professor at the Institute of Science Tokyo.

mance in diverse environments with varying lighting, weather, and camera setups. This will help assess its practical applicability and identify areas for improvement.

We believe our system's design has the potential to be adapted to a wide range of species, for example, giraffes and Valais Blacknose sheep, among others. We will construct other datasets for different animal species and apply our system to other species. This work leads to the widespread adoption of camera-based wildlife monitoring techniques across diverse ecological contexts.

Appendix

We thank and honor the Ueno Zoological Gardens²⁴⁾ for allowing us to capture HS images. This work was supported by JSPS KAKENHI 22K12081.

References

- 1) Bishop, C. M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 1 edition (2007).
- 2) Bishop, C. M. and Bishop, H.: *Deep learning: Foundations and concepts*, Springer Nature, Cham, Switzerland (2023).
- 3) Bradski, G.: The OpenCV Library (2000).
- 4) Burghardt, T., Thomas, B., Barham, P. J. and Calic, J.: Automated visual recognition of individual African penguins, *Fifth International Penguin Conference* (2004).
- 5) Dodd Jr, C. K.: Infrared reflectance in chameleons (Chamaeleonidae) from Kenya (1981).
- 6) Eba Japan Co.Ltd.: Hyper spectral camera NH9, <https://ebajapan.jp/products/hyper-spectral-camera/>.
- 7) Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M. and Grieve, B.: Towards on-farm pig face recognition using convolutional neural networks, *Computers in Industry*, Vol. 98, pp. 145–152 (2018).
- 8) Jocher, G.: Ultralytics YOLOv5 (2020).
- 9) Kingma, D. P.: Adam: A method for stochastic optimization (2014).
- 10) Kolmann, M., Kalacska, M., Lucanus, O., Sousa, L., Wainwright, D., Arroyo-Mora, J. and Andrade, M.: Hyperspectral data as a biodiversity screening tool can differentiate among diverse Neotropical fishes, *Scientific reports*, Vol. 11, No. 1, p. 16157 (2021).
- 11) Kumar, S. and Singh, S. K.: Visual animal biometrics: survey, *Iet Biometrics*, Vol. 6, No. 3, pp. 139–156 (2017).
- 12) Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324 (1998).
- 13) Magnusson, M., Sigurdsson, J., Armansson, S. E., Ulfarsson, M. O., Deborah, H. and Sveinsson, J. R.: Creating RGB Images from Hyperspectral Images Using a Color Matching Function, *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2045–2048 (2020).
- 14) Matkowsky, W. M., Kong, A. W. K., Su, H., Chen, P., Hou, R. and Zhang, Z.: Giant Panda Face Recognition Using Small Dataset, *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE (2019).
- 15) Matkowsky, W. M., Kong, A. W. K., Su, H., Chen, P., Hou, R. and Zhang, Z.: Giant panda face recognition using small dataset, *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 1680–1684 (2019).
- 16) Mielewicz, M., Liebis, F., Walter, A. and Greven, H.: Near-infrared (NIR)-reflectance in insects-Phenetic studies of 181 species, *Entomologie heute*, Vol. 24, pp. 183–215 (2012).
- 17) Noboru, Y., Ozasa, Y. and Tanaka, M.: Hyperspectral Image Dataset for Individual Penguin Identification, *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 9383–9387 (2024).
- 18) Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jovic, N. and Clune, J.: A deep active learning system for species identification and counting in camera trap images, *Methods in ecology and evolution*, Vol. 12, No. 1, pp. 150–161 (2021).
- 19) Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimeshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., pp. 8024–8035 (2019).
- 20) Pinto, F., Mielewicz, M., Liebis, F., Walter, A., Greven, H. and Rascher, U.: Non-invasive measurement of frog skin reflectivity in high spatial resolution using a dual hyperspectral approach, *PLoS One*, Vol. 8, No. 9, p. e73234 (2013).
- 21) Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., Matsuzawa, T., Biro, D. and Carvalho, S.: Chimpanzee face recognition from videos in the wild using deep learning, *Science Advances*, Vol. 5, No. 9, p. eaaw0736 (2019).
- 22) Schwalm, P. A., Starrett, P. H. and McDiarmid, R. W.: Infrared reflectance in leaf-sitting neotropical frogs, *Science*, Vol. 196, No. 4295, pp. 1225–1226 (1977).
- 23) Sherley, R., Burghardt, T., Barham, P., Cuthill, I. and Campbell, N.: Spotting the difference: towards fully-automated population monitoring of African penguins *Spheniscus demersus*, *Endangered Species Research*, Vol. 11, No. 2, pp. 101 – 111 (2010).
- 24) Tokyo Zoological Park Society: Ueno Zoological Gardens, <https://www.tokyo-zoo.net/english/ueno/index.html>.
- 25) Wu, J., Xu, W., He, J. and Lan, M.: YOLO for Penguin Detection and Counting Based on Remote Sensing Images, *Remote Sensing*, Vol. 15, No. 10, p. 2598 (2023).



Youta Noboru Youta Noboru received the B.Eng. degree in Information System Engineering from Tokyo Denki University in 2024. He is currently pursuing the Master degree in Machine Learning and Computer Vision at Tokyo Denki University. His research interests include visual perception, and hyperspectral sensing.



Yuko Ozasa Yuko Ozasa received her PhD degree in engineering from Kobe University in 2015. She was a postdoctoral researcher at National Institute of Advanced Industrial Science and Technology (AIST) in 2015. She had been a research associate at Graduate School of Science and Technology, Keio University from 2016 to 2020. She had been an assistant professor at Graduate School of System Design and Technology, Tokyo Denki University from 2020 to 2022. Since 2023, she has been an associate professor at the university. She has also been a guest professor at Gakushuin University since 2024. Her research interests include object recognition and grounding, visual perception, and hyperspectral sensing.



Masayuki Tanaka Masayuki Tanaka received his bachelor's and master's degrees in control engineering and a doctoral degree from Tokyo Institute of Technology in 1998, 2000, and 2003. He was a software engineer at Agilent Technologies from 2003 to 2004. He was a Research Scientist at Tokyo Institute of Technology from 2004 to 2008. He was an Associate Professor at the Graduate School of Science and Engineering, Tokyo Institute of Technology, from 2008 to 2016. He was a Visiting Scholar at Department of Psychology, Stanford University, from 2013 to 2014. He was an Associate Professor at School of Engineering, Tokyo Institute of Technology, from 2016 to 2017. He was a Senior Researcher at National Institute of Advanced Industrial Science and Technology from 2017 to 2020. He was an Associate Professor at School of Engineering, Tokyo Institute of Technology, from 2020 to 2023. Since 2023, he has been a Professor at Graduate Major in Engineering Sciences and Design, Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology. In 2024, following the university's integration, he became a Professor at the Institute of Science Tokyo.