CNNで抽出したパケットの特徴に基づくネットワークの異常検出

鷺坂 典雅[1,a)]　青木 茂樹[1,b)]　宮本 貴朗[1]

概要：近年，サイバー攻撃の増加に伴い，ネットワーク上の不正な通信を検出する侵入検知システム (IDS : Intrusion Detection System) の研究が盛んに行われている．一方，深層学習技術は画像認識分野において著しい発展を遂げており，特に大規模物体認識データセットを用いて学習させた CNN (Convolutional Neural Network) の中間層から抽出される特徴を，様々な分野に応用する研究が盛んに行われている．本稿では，画像処理分野で高い精度を示している CNN である VGG16 を特徴抽出器として用いて抽出した特徴に基づいて，ネットワークトラフィック中の異常トラフィックを検出する異常検知手法を提案する．まず，トラフィックデータを CNN に入力するために画像形式に変換する．この際，多様な攻撃に対応できるように複数の手法を適用して画像に変換する．次に，生成した画像を VGG16 に入力し，VGG16 の中間層から特徴ベクトルを抽出する．抽出した特徴ベクトルを識別するためのニューラルネットワークを学習用データにより学習させ，学習用データとは別に取得した識別用データに含まれる異常トラフィックを検出する．CICIDS 2017 データセットを用いて実験を実施し，本手法の有効性を確認した．

キーワード：ネットワークの異常検出，CNN，VGG16，IDS

Network anomaly detection based on features of packets extracted by CNN

Norimasa sagisaka[1,a)]　Shigeki Aoki[1,b)]　Takao Miyamoto[1]

Abstract: In recent years, with increase of cyber-attacks, research on IDS (Intrusion Detection System) to detect unauthorized communication has been actively conducted. On the other hand, deep learning technology has made remarkable progress in field of image processing. Thus, research on detecting cyber-attacks in network traffic by applying deep learning, which has shown high performance in the image processing field, to IDS is also being actively conducted. In this paper, we propose a method for detecting and identifying cyber-attacks with high accuracy using VGG16, which is a CNN that has shown high accuracy in the field of image processing. We use VGG16 as a feature extractor, and extract features from traffic data by VGG16. We learn the features by neural network and detect cyber-attacks. Experiments are conducted on CICIDS 2017 dataset to confirm the effectiveness of the proposed method.

Keywords: Network Anomaly Detection，CNN，VGG16，IDS

1. はじめに

近年のサイバー攻撃の増加に伴い，ネットワーク上の不正なトラフィックを検出する侵入検知システム (IDS: Intrusion Detection System) の研究が盛んに行われている．IDS はシグネチャ型とアノマリ型の 2 種類に分けら

[1] 大阪府立大学大学院人間社会システム科学研究科
Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University
[a)] saa01119@edu.osakafu-u.ac.jp
[b)] aoki@kis.osakafu-u.ac.jp

れる．代表的なシグネチャ型 IDS として，Snort [1] や Suricata [2]，Zeek [3] 等が挙げられる．シグネチャ型 IDS はパターンファイルに基づいて異常を検出する方式であり，誤検知は少ないがパターンファイルに定義されていない攻撃は検出できないという欠点がある．一方，代表的なアノマリ型 IDS として，文献 [4–7] の手法が挙げられる．アノマリ型 IDS は正常な通信のみを含むデータを正常状態と定義し，そこから外れた状態を異常として検出する方式である．これにより，シグネチャ型 IDS の問題点であった未知の異常の検出が可能となる．しかし，シグネチャ型 IDS に比べて誤検知が多く発生することや，異常の発生自体は検知できるものの，どのような異常であるかを識別できない欠点がある．

一方，深層学習技術が画像処理分野において著しい発展を遂げている．そして，画像処理分野で高い性能を示している深層学習を IDS に応用することで，ネットワークトラフィック中の異常を検知する研究 [8–10] も盛んに行われている．これらの研究では，パケットの特徴を画像形式で表現して深層学習手法で学習し，異常を識別している．

また，大規模物体認識データセットを用いて学習させた畳み込みニューラルネットワーク (CNN：Convolutional Neural Network) の中間層から抽出される特徴ベクトルの汎用性が近年注目されており，文献 [11] では CNN の中間層から抽出した特徴を，物体認識や画像検索，カテゴリ識別など様々な分野に応用した研究事例を紹介している．

本研究では，高精度な IDS の構築のために，高い識別能力を持つ大規模物体認識データセットで学習済みの CNN に注目し， CNN の中間層から抽出したパケットの特徴をニューラルネットワークで学習・識別する手法を提案する．まず，トラフィックデータを画像認識手法である CNN に入力するため，画像形式のデータに変換する．次に，攻撃ごとにトラフィックの特徴が異なるため，画像化の方法によって検出できる攻撃が異なると考え，複数の手法で画像に変換する．変換した画像を，大規模データセットで学習を行った CNN に入力し，CNN の中間層からの出力を，トラフィックデータの特徴ベクトルとして抽出する．その後，抽出した特徴ベクトルをニューラルネットワークで学習し，新たな通信中に含まれる攻撃通信を種類ごとに識別する．

以下，2 節で，関連研究について述べ，3 節では提案手法について説明する．4 節では実験条件と実験結果，考察について述べ，5 節でまとめと今後の課題について述べる．

2. 関連研究

本研究に関連する研究として，深層学習技術のネットワークの異常検知分野への応用に関する文献 [8–10] ついて述べる．

画像処理分野で高い性能を示している CNN を IDS に応用する研究として，文献 [8] では 1 パケットごとにトラフィックデータを画像に変換し，CNN を用いてトラフィックデータを学習することで，不審なパケットを検出し，攻撃に関する情報を抽出する手法が提案されている．この手法では 1 パケットごとに画像に変換しているため，DoS 攻撃のような連続したパケットに特徴が現れる攻撃を検出できないことが課題として挙げられている．

文献 [9, 10] では，深層学習の中でも教師データを必要としない敵対的生成ネットワーク (GAN:Generative Adversarial Network) に CNN を組み込むことで，精巧な画像を生成することができる DCGAN(Deep Convolutional Generative Adversarial Network) を用いてアノマリ型 IDS を構築している．これらの手法では，DCGAN は学習した画像と類似する画像は生成できるが，学習していない画像に類似する画像は生成できないという性質を利用することで，ネットワークトラフィックの異常を検出している．まず，正常な通信のトラフィックデータから変換した画像を DCGAN で学習する．その後，テスト用のトラフィックデータを学習時と同様にテスト用画像へと変換し，変換した画像に類似する画像 を DCGAN を用いて生成する．そして，生成した画像とテスト用画像との類似度から算出した異常度を基に，正常な通信と異常な通信を識別している．文献 [9, 10] では，複数のパケットを組み合わせて 1 枚の画像を生成することにより，文献 [8] の課題であった連続したパケットに特徴が現れる攻撃を検出している．また，文献 [10] では，複数の画像化手法を用いることで，より多くの異常パターンに対応し，検知精度を向上させている．一方，正常通信との類似度によって異常を検出しているため，検出された異常の種類までは識別することができていない．

3. 提案手法

本研究では，画像処理分野で高い性能を示している CNN を用い，その中間層から抽出される特徴ベクトルをニューラルネットワークで学習することで，高い識別能力を有する識別機を構成し，高精度に異常を識別する手法を提案する．本手法は学習と異常検出の 2 つのプロセスで構成されている．学習プロセスではまず，あるネットワークで送受信されるパケットを pcap 形式でキャプチャし，トラフィックデータとする．キャプチャしたトラフィックデータは，そのままの形式では CNN に入力出来ないため，入力可能な画像形式のデータに変換する．ここで，画像化の方法によって検出できる攻撃が異なると考え，文献 [10] を参考に 3 種類の手法で画像に変換する．次に，大規模データセットで学習を行った CNN に画像を入力し，CNN の中間層から得られる出力を，各画像のトラフィックデータを表す特徴ベクトルとして抽出する．そして，教師データとして，各特徴ベクトルに攻撃の種類を示すラベルを付与

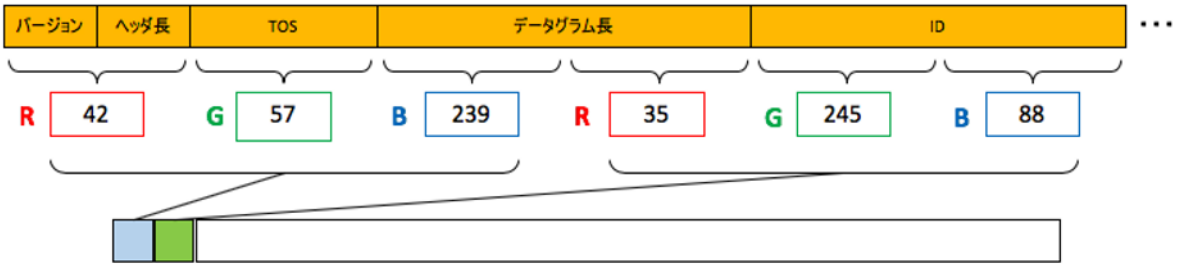

図 1 画像への変換の概要
Fig. 1 Overview of image conversion method.

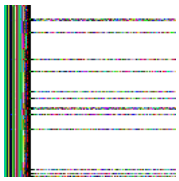

図 2 手法 (1) の正常画像の例
Fig. 2 Examples of Normal image (method1).

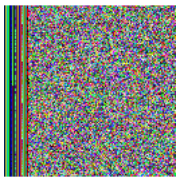

図 3 手法 (2) の正常画像の例
Fig. 3 Examples of Normal image (method2).

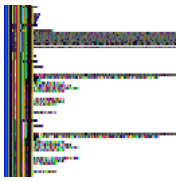

図 4 手法 (3) の正常画像の例
Fig. 4 Examples of Normal image (method3).

する．その後，抽出した特徴ベクトルと教師データを攻撃識別用ニューラルネットワークで学習する．

異常検出プロセスでは，学習プロセスと同様にトラフィックデータを画像化し，CNN の中間層から特徴ベクトルを抽出する．抽出した特徴ベクトルを，攻撃識別用ニューラルネットワークに入力してトラフィックに含まれる攻撃の種類を識別する．

3.1 トラフィックデータの画像変換

本研究では検知対象とする攻撃の特徴に合わせて，3 種類の画像化手法により，収集したトラフィックデータを画像に変換する．

（1） 宛先 IP アドレスが同一のパケットを抽出し画像化

- この手法は，DDoS 攻撃や Port Scan 攻撃など特定のホストに対する攻撃の検出を想定している．

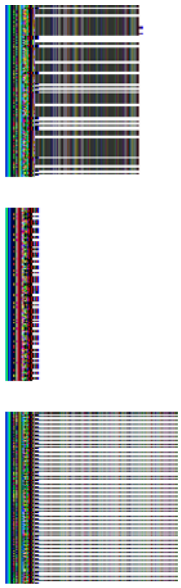

図 5 異常画像の例
Fig. 5 Examples of Anormal image.

（2） 宛先 IP アドレス / 宛先ポート番号の組み合わせが同一のパケットを抽出し画像化

- XSS 攻撃や，Brute Force 攻撃など特定のサービスに対する攻撃の検出を想定した画像化手法である．これらの攻撃は，通信内容に関する攻撃であるため，パケットのペイロード部分のみでの画像化手法も考えられるが，ここでは，攻撃対象のホストを特定する際にヘッダ情報が必要であると考え，ヘッダ部分も使用している．

（3） 宛先 / 送信元 IP アドレスの組み合わせが同一のパケットを抽出し画像化

- この手法は，DoS 攻撃などの特定のホストから特定のホストに対する攻撃の検出を想定した画像化手法である

それぞれの手法で収集したトラフィックデータを，図 1 に示す方法により画像に変換する．抽出したトラフィックデータから 1 パケットずつ取得し，パケットの先頭から 128 × 3 バイトを 8 bit 単位で 0~255 の数値として読み

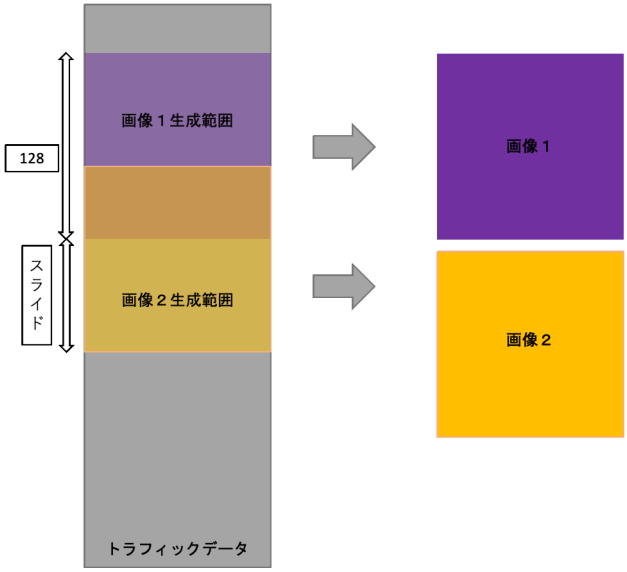

図 6 生成範囲のスライドによる画像生成手法
Fig. 6 Image generation method that shift generation range.

込む．そして変換した数値を 1 画素あたり RGB の 3 つの値に割り当てることで，1 × 128 画素の画像 1 枚を作成する．これを 128 パケットに対して行い，128 × 128 画素の画像 1 枚を作成する．このプロセスを繰り返すことで，トラフィックデータを画像に変換する．以上の処理により，生成した正常画像の例を図 2 3 4 に，異常画像の例を図 5 に示す．図中の 1 行が 1 パケットを表しており，画像全体で 128 パケットが表されている．また，白色で表現されている部分はデータが含まれていない部分を表す．このようにトラフィックデータを画像に変換することでトラフィックの特徴を顕著に表すことができる．

生成される画像枚数が少ない場合，学習プロセスの際に過学習が発生しやすくなる．そのため，画像生成範囲のスライドを行うことで，生成する画像の枚数を増加させる．画像生成範囲のスライドを用いた画像の増加方法を図 6 に示す．この方法では，画像生成に利用するトラフィックデータの範囲をスライドさせることで，一度画像に変換したパケットデータを再度他の画像に使用し，トラフィックデータの量が少ない場合でも多数の画像を生成している．

3.2 特徴ベクトルの抽出

変換した画像から，特徴ベクトルを抽出する．本研究では，大規模物体認識データセットで学習済みの CNN である VGG16 [12] を適用することで画像化したトラフィックデータに現れる特徴を抽出する．VGG16 は，画像認識技術に関するコンペティションである ILSVRC-2014 において 2 位の成績を獲得した CNN で，物体認識分野で高い性能を発揮しており 1000 クラスの画像を分類できる．VGG16 は畳み込み層が 13 層，全結合層が 3 層，計 16 層から構成され，ImageNet [13] を用いた事前学習によって

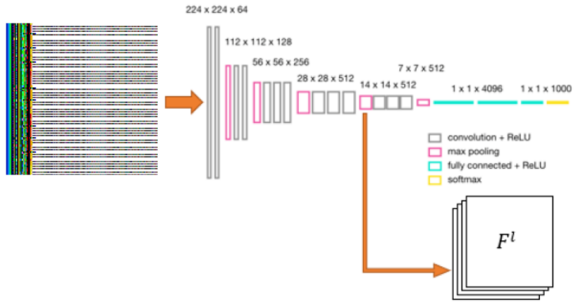

図 7 VGG16 による特徴行列抽出の概要
Fig. 7 Overview of extracting feature map by VGG16.

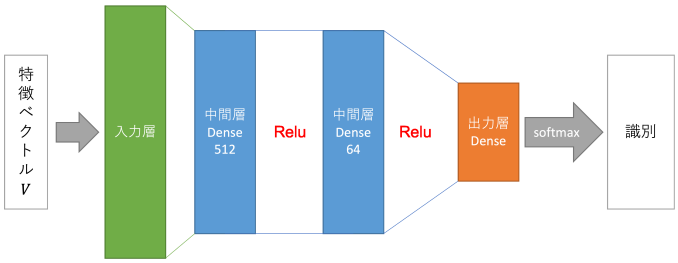

図 8 ニューラルネットワークの構成
Fig. 8 Configuration of Neural network .

転移学習を容易に行えることも特徴の一つである．

図 7 に示すように変換した画像を 1 枚 VGG16 に入力し，適当な中間層 l から特徴行列 $\boldsymbol{F}^l \in \mathbb{R}^n$ を抽出する．抽出した特徴行列 $\boldsymbol{F}^l$ から，式 (1) により特徴ベクトル $\boldsymbol{V}$ を算出する．

$$\boldsymbol{V} = \left(\boldsymbol{F}^l_{1,1}, \boldsymbol{F}^l_{1,2}, \cdots, \boldsymbol{F}^l_{1,n}, \boldsymbol{F}^l_{2,2}, \cdots, \boldsymbol{F}^l_{n,n}\right) \qquad (1)$$

3.3 攻撃識別用ニューラルネットワークの学習と攻撃の識別

VGG16 の中間層から抽出した特徴ベクトル $\boldsymbol{V}$ を入力値，攻撃の種類を教師データとして，攻撃識別用ニューラルネットワークを学習する．ニューラルネットワークは Relu 関数を用いた中間層が 2 層，出力層が softmax 関数の計 3 層で構成する．ニューラルネットワークの構成を図 8 に示す．

Relu 関数とは，関数への入力値が 0 以下の場合には出力値が 0 となり，入力値が 0 より大きい場合には出力値が入力値と同じ値となる活性化関数である．Relu 関数を式 (2) に示す．

$$f(x) = \begin{cases} x & (x \geq 0) \\ 0 & (\text{ otherwise }) \end{cases} \qquad (2)$$

また，softmax 関数とは，多クラス分類モデルで利用され

る活性化関数の 1 つであり，入力データがそのクラスに属する確率を出力する．softmax 関数を式 (3) に示す．

$$y_i = \frac{\exp(x_i)}{\sum_{k=1}^{n} \exp(x_k)} \tag{3}$$

さらに，過学習を防ぐため学習の際にドロップアウトを行う．ドロップアウトとは，学習の際にネットワーク内のいくつかのノードを無効化する手法である．無効化する割合は学習前に設定する．

識別を行う際は，テスト用トラフィックデータを 3.1 節で述べた 3 つの手法で画像に変換して，学習時と同様に 3.2 節で述べた方法で特徴ベクトル $\boldsymbol{V}$ を抽出し，攻撃識別用ニューラルネットワークに入力する．そして，ニューラルネットワークから出力されるクラスの攻撃が，画像中のパケットに含まれると識別する．

4. 実験と考察

本手法の有効性を確認するために実験を行った．実験には CICIDS2017 データセット [14] を使用した．評価指標には loss・accuracy 及び，Recall (再現率) と Precision (適合率) を用いた．loss は出力値と教師データの与える正解との誤差を示したものであり，accuracy は教師データに対する出力値の正答率を示している．また，Recall とは，異常全体のうち，異常であると正しく識別された割合を表す．Recall の算出式を式 (4) に示す．

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

Precision とは，異常と識別されたもののうち，実際に異常であった割合を表す．Precision の算出式を式 (5) に示す．

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

ここで，TP (True Positive) は異常を正しく異常と識別した数，FP (False Positive) は正常を異常と誤識別した数，TN (True Negative) は正常を正しく正常と識別した数，FN (False Negative) は異常を正常と誤識別した数を表す．

4.1 実験条件

4.1.1 実験データセット

CICIDS2017 データセットは, 攻撃者端末から組織内ネットワークへのサイバー攻撃を想定した研究用データセットである．データセットは 2017/7/3 (月) から 2017/7/7 (金) までのトラフィックをキャプチャしており，月曜日は正常通信のみ，それ以外の日は攻撃通信と正常通信の両方を含むトラフィックである．含まれている攻撃は，水曜日に DoS 攻撃 (slow loris / SlowHTTPTest / GoldenEye / Hulk)，木曜日に Brute Force 攻撃と XSS 攻撃，金曜日には Port scan 攻撃，DDoS 攻撃のトラフィックがそれぞれ含まれている．

表 1 データセットの詳細
Table 1 Details of dataset.

	学習用データ	テスト用データ
slow loris	2017/07/05(水) 9:48:46-9:58:46	2017/07/05(水) 9:58:46-10:08:46
SlowHTTPTest	2017/07/05(水) 10:15:38-10:25:38	2017/07/05(水) 10:25:38-10:35:38
Hulk	2017/07/05(水) 10:43:24-10:53:24	2017/07/05(水) 10:53:24-11:03:24
GoldenEye	2017/07/05(水) 11:10:24-11:20:04	2017/07/05(水) 11:20:04-11:23:04
Brute Force	2017/07/06(木) 9:31:10-9:41:10	2017/07/06(木) 9:41:10-9:51:10
XSS	2017/07/06(木) 10:15:35-10:25:35	2017/07/06(木) 10:25:35-10:35:35
Port scan	2017/07/07(金) 14:55:00-18:05:00	2017/07/07(金) 15:05:00-15:15:00
DDoS	2017/07/07(金) 15:57:00-16:07:00	2017/07/07(金) 16:07:00-16:16:00

今回の実験では，学習用データ・テスト用データに正常通信と異常通信を含む区間から 7 種類の攻撃毎にそれぞれ 10 分ずつ用いた．ただし，GoldenEye 攻撃と DDoS 攻撃を含むトラフィックデータに関しては，攻撃が行われた時間が 20 分未満であったため，学習用として 10 分，テスト用に残り区間を用いた．表 1 にそれぞれの攻撃の学習用データとテスト用データの区間の詳細を示す．

4.2 ラベル付与

学習用データにラベルを付与した．ラベル付与の概要を図 9 に示す．攻撃元との通信パケットを異常パケットとして，3.1 節で述べた手法を用いてトラフィックデータから画像を生成し，異常パケットが含まれている画像を異常画像とした．また，異常画像には含まれている攻撃の種類に応じてそれぞれのラベルを付与し，正常パケットのみから生成されている画像には正常ラベルを付与した．

4.2.1 特徴抽出と学習

特徴ベクトル $\boldsymbol{V}$ は，第 5 層のプーリング層 ($7 \times 7 \times 512$) から抽出した特徴行列 $\boldsymbol{F}_k^5$ から算出した．

攻撃識別用ニューラルネットワークは，エポック数 20, バッチサイズ 128, 学習率 0.001，ドロップアウトは 0.25 として学習を行った．また，過学習を防ぐために，学習が収束した時に自動的に学習を止める Early Stopping を使用した．

4.3 実験結果と考察

3 種類の画像化手法が，それぞれどのような攻撃の識別に対して有効であるかを確認するために識別実験を行っ

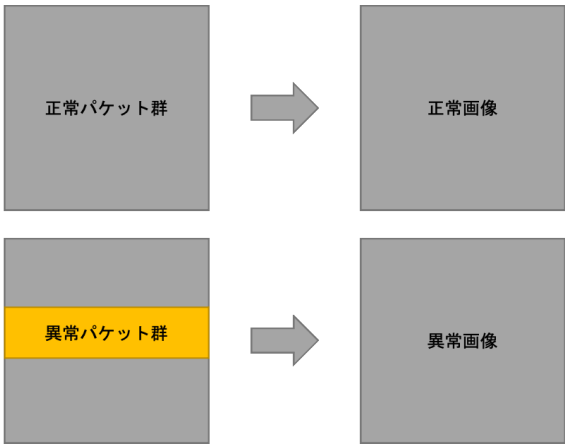

図 9 ラベル付与の概要

Fig. 9 Overview of Labeling.

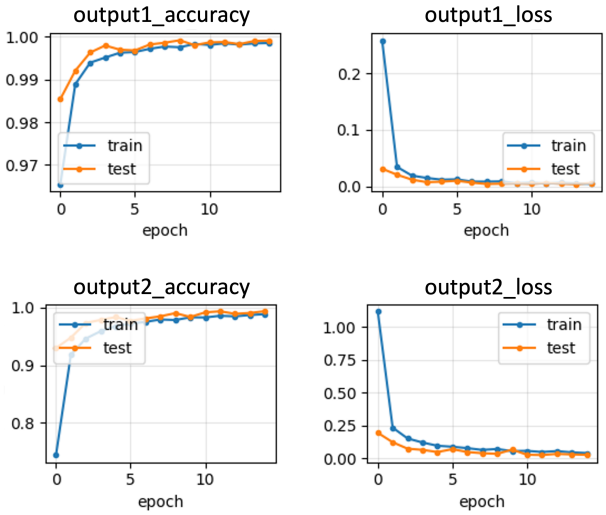

図 10 手法 (1) の結果

Fig. 10 Overview of Anomaly Detection (method1).

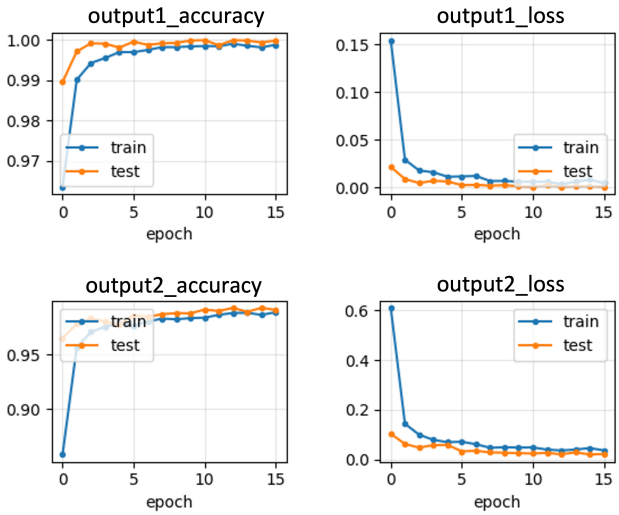

図 11 手法 (2) の結果

Fig. 11 Overview of Anomaly Detection (method2).

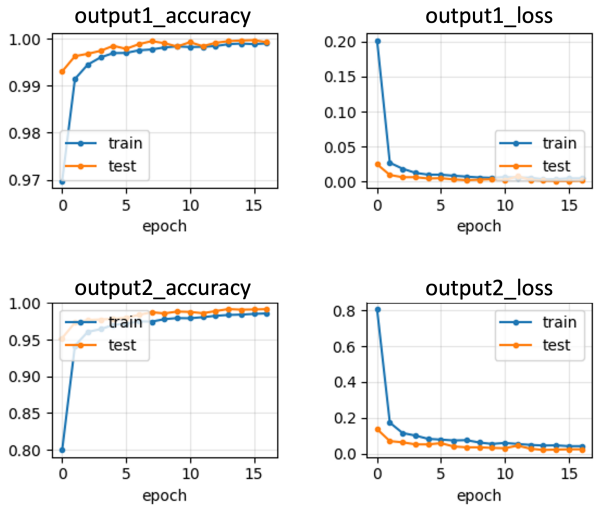

図 12 手法 (3) の結果

Fig. 12 Overview of Anomaly Detection (method3).

た．また，正常と異常のみ識別する実験と，攻撃の種類を識別する実験の 2 通りの実験を行った．

図 10，11，12 は学習の際の各エポックに対するトレーニングデータ (train) とバリデーションデータ (test) の loss，accuracy のスコアを示したグラフであり，横軸がエポック数，縦軸がスコアを表している．また，output1 は正常か異常かの 2 値分類，output2 は攻撃の種類ごとの識別の学習結果を示している．表 2 は最終エポックにおけるテストデータの loss と accuracy を示している．

また，学習用データで学習したニューラルネットワークを用いて，テスト用データを識別した結果を Recall，Precision によって評価した．2 値分類の識別結果を表 3，攻撃の種類ごとの識別結果を表 4 にそれぞれ示す．

図 10，11，12 の結果から，5 エポック程度で収束しており，表 2 の output1(異常と正常の識別) では，全ての画像化手法で loss が 0.006 以下，accracy は 0.999 以上となり，また，output2(攻撃の種類の識別) では，2 値分類と比較するとやや劣るものの，各画像化手法で loss は 0.025 以下，accracy は 0.991 以上と高い精度で識別できることを確認した．また，表 3 の結果から，正常と異常の 2 値分類においては各画像化手法で Recall・Precision 共に 98.5% を超えており，手法 (3) では 99% 以上と高い精度で識別できることを確認した．

次に，攻撃の種類ごとの識別実験を行なった．表 4 の結果から，DDoS 攻撃と正常のトラフィックデータに対しては 各画像化手法で Recall・Precision 共に 99% 以上の精度で識別できることを確認した．また，Hulk 攻撃，Brute Force 攻撃，XSS 攻撃に関しては手法 (1)，手法 (2) において Recall，Precision 共に 96.2% 以上の精度で識別できている．さらに，手法 (1) の XSS 攻撃と手法 (2) の Brute Force 攻撃の識別精度は 100% だった．一方で，slow loris

表 2　最終エポックにおける loss, accuracy
Table 2　Scores of test loss and test accuracy in a final epoch.

	手法 (1)		手法 (2)		手法 (3)	
	loss	accuracy	loss	accuracy	loss	accuracy
output1	0.006	0.999	0.001	0.999	0.002	0.999
output2	0.025	0.994	0.022	0.991	0.024	0.991

攻撃，SlowHTTPTest 攻撃に関しては，Precision は少し低下するものの，手法 (2) では Recall が 92% 以上となっている．Port scan 攻撃は各画像化手法で 97.7% 以上の高い Precision を得ることができたが Recall は 44.6% 以下と低く，特に手法 (2) では 4% となった．DoS-GoldenEye 攻撃は各画像化手法で Recall, Precision 共に低くうまく識別することができなかった．

文献 [8] では主に Port scan 攻撃を対象としていたが，本手法では Hulk 攻撃，Brute Force 攻撃，XSS 攻撃，DDoS 攻撃に対して，高い識別精度であることを確認した．一方，Port scan 攻撃，Hulk 以外の DoS 攻撃を含むトラフィックデータでは他の攻撃に対する識別精度よりも低い結果となった．これに関して，多くの Port scan 攻撃，GoldenEye 攻撃が slow loris 攻撃，SlowHTTPTest 攻撃に誤識別されていた．これは，今回の実験で使用したデータに含まれる Port scan 攻撃で用いられていた手法が学習用データとテスト用データで異なっていた為であると考えられる．更に，Port scan 攻撃の異常パケットはパケットサイズが小さい為，生成された画像に含まれる空白部分の割合が大きいことも要因であると考えられる．また，Golden Eye 攻撃に関しては，学習用データセットに含まれるパケット数が少なく画像生成範囲のスライド数を非常に小さくすることで，学習用画像を増加させていた．そのため，学習された特徴に偏りがあったと考えられる．この問題は，学習パターンを増加させることで改善できると考えられる．一方，手法 (3) では Hulk 攻撃が全て Brute Force 攻撃に，XSS 攻撃は全て DDoS 攻撃に誤って識別されていた．学習率を 0.0001 に変更し，再度実験を行なったところ XSS 攻撃は Recall，Precision 共に 1 と正確に識別されることを確認した．その為，XSS 攻撃に関しては学習が不十分であり，適切な学習率を設定することにより改善が可能であると考えられる．しかし， Hulk 攻撃は学習率を変更した場合も Brute Force 攻撃に誤識別されており，これは手法 (3) で抽出したパケットに現れる Hulk 攻撃と Brute Force 攻撃の特徴が酷似している為だと考えられる．

表 4 に示す結果から，攻撃ごとに最も高い識別結果となる画像化手法は異なり，適切な画像化手法を選択することができれば表 4 の下線で示した精度の Recall ，Precision を得られることが確認できる．今後，画像化手法をさらに増加させると共に，それぞれの画像化手法での識別結果の適切な組み合わせ方を検討することによって，高精度な IDS を構築したいと考えている．

表 3　2 値分類実験結果
Table 3　Experimental Result of Binary Classification.

	手法 (1)		手法 (2)		手法 (3)	
	Recall	Precision	Recall	Precision	Recall	Precision
2 値分類	0.998	0.986	0.985	0.998	0.996	0.992

表 4　多クラス分類実験結果
Table 4　Experomental Result of Multiclass Classification.

	手法 (1)		手法 (2)		手法 (3)	
	Recall	Precision	Recall	Precision	Recall	Precision
slow loris	0.882	0.647	0.923	0.834	0.914	0.655
SlowHTTPTest	0.822	0.522	0.946	0.734	0.776	0.814
GoldenEye	0.455	0.167	0.5	0.077	0.25	0.042
Hulk	0.997	0.999	0.995	1	0	0
Brute Force	1	0.986	1	1	1	0.5
XSS	1	1	1	0.962	0	0
Port scan	0.446	0.977	0.04	1	0.352	0.985
DDoS	0.998	0.994	1	0.998	0.992	0.996
Normal	0.913	0.979	0.993	0.930	0.981	0.983

5. おわりに

本研究では，パケットを画像変換し，VGG16 の中間層から抽出した特徴を基にニューラルネットワークを学習させることでネットワーク中の異常を識別する手法を提案した．実験では，CICIDS2017 データセットを用いて，3 種類の画像化手法でトラフィックデータから画像を生成した．生成した画像を正常と異常に識別する実験と，攻撃の種類ごとに識別する実験により有効性を確認した．実験の結果，用いた全ての画像化手法で正常画像と異常画像を高精度に識別できることを確認できた．一方，攻撃の種類ごとの識別に関しては，画像化の手法によって攻撃ごとに識別精度が異なり，それぞれの攻撃に最適な画像化手法が存在することを確認した．

今後の課題としては，ニューラルネットワークのパラメータの調節や，今回の手法では検知精度が低かった攻撃に対する適切な画像化手法の検討が挙げられる．また，画像化手法ごとの識別結果の組み合わせ手法の検討を行い，高精度な IDS を構築したいと考えている．

参考文献

[1] Snort，<https://www.snort.org/>(参照 2021-08-20).
[2] Suricata，<https://suricata.io/>(参照 2021-08-20).
[3] Zeek，<https://zeek.org/>(参照 2021-08-20).
[4] 小島俊輔，中嶋卓雄，末吉敏則: エントロピーベースのマハラノビス距離による高速な異常検知手法，情処学論，vol.52，no.2，pp.656-668 (2011).
[5] 佐藤陽平，和泉勇治，根元義章: 複数の検出モジュールによるネットワーク異常検出の高精度化，信学技報，NS2004-144，pp.45-48 (2004).
[6] 平松尚利，和泉勇治，角田裕: 複数の通常状態を用いたネットワーク異常検出，信学技報，CS2006-32，pp.61-66 (2006).
[7] 鷺坂典雅，青木茂樹，宮本貴朗: トラフィックデータのエントロピーに基づくアノマリ型 IDS の構築，コンピュー

タセキュリティシンポジウム 2020 論文集，pp.1033-1039 (2020).
[8] 池端悠人，木村航佑，松尾美咲，加藤雅彦: CNN を用いたネットワークトラフィック異常検知と異常トラフィックからの脅威情報抽出，コンピュータセキュリティシンポジウム 2019 論文集，pp.1108-1115 (2019).
[9] 日置裕士，青木茂樹，宮本貴朗: DCGAN を用いたネットワークトラフィッ クの異常検出，コンピュータセキュリティシンポジウム 2018 論文集, 2C2-1, pp.341-347 (2018).
[10] 森岡卓哉，青木茂樹，宮本貴朗: DCGAN とパーティクルフィルタを用いたネットワークトラフィック異常検出，情処研報，2021-CSED-92(61)，pp.1-8 (2021).
[11] 中山英樹: 深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習，信学技報，sp2015-45, pp. 55-59 (2015).
[12] Keran Simonyan, Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition, Proc. of ICLR2015 (2015).
[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. Proc. of CVPR2009, pp.2-9 (2009).
[14] I. Sharafaldin, A. Habibi Lashkari and A. A. Ghorbani: Toward generating a new intrusion detection dataset and intrusion traffic characterization, Proc. of 4th ICISSP, pp. 108-116 (2018).