The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

# 図形認識のための多層ニューラルネットにおける大局構造の抽出

Detecting community structure in layered neural networks for diagram recognition

渡邊千紘 *1 Chihiro Watanabe　平松薫 *1 Kaoru Hiramatsu　柏野邦夫 *1 Kunio Kashino

*1NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

Layered neural networks (LNNs) have realized high recognition performance for various real datasets, however, it is difficult for human beings to understand their training results. Conventionally, we have proposed network analysis methods for extracting simplified structure of a trained LNN, by detecting communities of units based on the similarity of connection patterns. In this work, we propose a new method for representing the community structure in a LNN, by using connection weights between pairs of communities. By experiment using the dataset of diagram recognition, we show that our new method provides clues for interpreting the roles of each community in a LNN, in terms of which community in input-side adjacent layer is the most important for it in prediction.

## 1. はじめに

多層ニューラルネットは高次元空間に存在する様々な実データにおいて，高い予測性能を実現している．しかし，その内部表現は非常に多くのパラメータが非線形かつ階層的に組み合わせられた形で構成され，人が理解することは難しい．我々はこれまで，データから学習された多層ニューラルネットに対し，ネットワーク解析を適用することにより，各層において隣接する層のユニットと似た結合パターンを持つユニットのグループ (コミュニティ) を推定する方法を提案してきた [渡邊 17a, Watanabe 18, Watanabe 17b, Watanabe 17c]．これにより，学習結果のネットワーク構造を単純化し，大局的に捉えた表現を獲得することが可能となり，この結果から，ニューラルネットの各部分が推論において果たす役割を推察することができるようになった．これらの研究においては，ニューラルネットから抽出された各コミュニティ間に存在する複数の結合を，閾値処理に基づいて 1 本の結合束として表現することにより，構造の単純化を行っていたが，コミュニティ間がどの程度強く結びついているかを知る方法は存在していなかった．

本研究では，多層ニューラルネットから抽出されたコミュニティ構造に対し，新たな結合束の定義法と可視化手法を適用することにより，各コミュニティが入力側の層におけるどのコミュニティと最も強く結びついているかを知ることを可能にした．また，実際に図形認識のデータセットを用いて学習した多層ニューラルネットに対し，提案法を適用することにより，各コミュニティの役割について考察を行った．

## 2. ニューラルネットのコミュニティ構造抽出

確率的最急降下法に基づき，データから多層ニューラルネットの学習を行う．この際，LASSO[Ishikawa 90, Tibshirani 94] と呼ばれる手法を適用することにより，冗長な結合重みの値をゼロに近づけ，疎なネットワークを獲得することができる．詳細なニューラルネットの学習方法は，既存研究 [渡邊 17a] と同じである．ただし，実験では，学習を安定させるために，各反復において学習データを確率的に選ぶのではなく，各クラスの学習データを 1 つずつ順に選ぶこととした．

連絡先: watanabe.chihiro@lab.ntt.co.jp

データから学習された多層ニューラルネットにおいて，結合重みの符号を区別した上で，似た結合パターンを持つユニットを分類するコミュニティ抽出法として，既存研究 [渡邊 17a] が存在する．これは，確率モデルに基づくコミュニティ抽出法 [Newman 07] を，多層ニューラルネット向けに拡張した手法であり，本研究でもこの手法 [渡邊 17a] を用いてニューラルネットの各層におけるコミュニティ構造を推定する．

多層ニューラルネットの各層において，隣接する層との間の結合関係を 4 つの隣接行列 $A^+ = \{A^+_{i,k}\}$, $A^- = \{A^-_{i,k}\}$, $B^+ = \{B^+_{k,j}\}$, $B^- = \{B^-_{k,j}\}$ で定義する．ただし，$k = 1, \cdots, l$. ここで，$A^+(A^-)$ の各要素 $A^+_{i,k}(A^-_{i,k})$ は，入力側の隣接層における $i$ 番目のユニットと，各層における $k$ 番目のユニットとの間にある結合重みが $\xi$ 以上 ($-\xi$ 以下) ならば 1, そうでなければ 0 として定義する．同様に，隣接行列 $B^+(B^-)$ を，各ユニットと，出力側の隣接の層における各ユニットとの間の結合重みから定義する．

各層で，コミュニティ$c$にユニットが属する確率を $\pi = \{\pi_c\}$, コミュニティ$c$のユニットに入力される正 (負) の結合の結合元が入力側の隣接層における $i$ 番目のユニットである確率を $\tau^+ = \{\tau^+_{c,i}\}(\tau^- = \{\tau^-_{c,i}\})$, コミュニティ$c$のユニットから出力される正 (負) の結合の結合先が出力側の隣接層における $j$ 番目のユニットである確率を $\tau'^+ = \{\tau'^+_{c,j}\}(\tau'^- = \{\tau'^-_{c,j}\})$ とおく．

隣接行列 $A^+, A^-, B^+, B^-$ の尤度を最大化するパラメータ $\pi, \tau^+, \tau^-, \tau'^+, \tau'^-$ は，以下の更新式を繰り返し計算することにより，局所解として求めることができる (EM 法)．ここで，$q_{k,c}$ は $k$ 番目のユニットがコミュニティ$c$に属する確率を表し，$k$ 番目のユニットが属するコミュニティは，EM 法の最終反復において $q_{k,c}$ を最大化するコミュニティ$c$として定義する．

$$q_{k,c} = \Big\{\pi_c \Big[\prod_i \Big(\tau^+_{c,i}\Big)^{A^+_{i,k}} \Big(\tau^-_{c,i}\Big)^{A^-_{i,k}}\Big] \Big[\prod_j \Big(\tau'^+_{c,j}\Big)^{B^+_{k,j}} \Big(\tau'^-_{c,j}\Big)^{B^-_{k,j}}\Big]\Big\} \Big/ \Big\{\sum_s \pi_s \Big[\prod_i \Big(\tau^+_{s,i}\Big)^{A^+_{i,k}} \Big(\tau^-_{s,i}\Big)^{A^-_{i,k}}\Big] \Big[\prod_j \Big(\tau'^+_{s,j}\Big)^{B^+_{k,j}} \Big(\tau'^-_{s,j}\Big)^{B^-_{k,j}}\Big]\Big\}, \quad \pi_c = \frac{\sum_k q_{k,c}}{l},$$

$$\tau^{+}_{c,i} = \frac{\sum_k q_{k,c} A^{+}_{i,k}}{\sum_{k,i} q_{k,c} A^{+}_{i,k}}, \quad \tau^{-}_{c,i} = \frac{\sum_k q_{k,c} A^{-}_{i,k}}{\sum_{k,i} q_{k,c} A^{-}_{i,k}},$$
$$\tau'^{+}_{c,j} = \frac{\sum_k q_{k,c} B^{+}_{k,j}}{\sum_{k,j} q_{k,c} B^{+}_{k,j}}, \quad \tau'^{-}_{c,j} = \frac{\sum_k q_{k,c} B^{-}_{k,j}}{\sum_{k,j} q_{k,c} B^{-}_{k,j}}.$$

## 3. 結合重みの値に基づく大局構造の定義

多層ニューラルネットから推定された各コミュニティ間に存在する複数の結合を，以下に示す方法に基づいて 1 本の結合束として表現する．まず，入力側の層のコミュニティ$c$ と出力側の層のコミュニティ$c'$ に対し，その間にある全ての結合について重みの絶対値の平均 $r_{c,c'}$ を結合束の重みとし，重み $r_{c,c'}$ に比例した太さの線で結合束を描画するものとする．ただし，結合束の太さは全ての層において最小値, 最大値が同じになるように正規化した．次に，結合束を疎にするために，各層間において，$r_{c_1,c_2}$ が $\min_{c'} \max_c r_{c,c'}$ 未満の場合はコミュニティ$c_1, c_2$ 間の結合束を削除するものとし，それ以外の場合はコミュニティ$c_1, c_2$ 間の結合束があるものとする．また，実験では，さらに視認性を向上するために，各出力側のコミュニティ$c'$ に対し，$r_{c,c'}$ の値を降順に並べたとき，1 番目 (2 番目) のコミュニティ$c$ からの結合束を青色 (黄土色) で描画した．

## 4. 実験

実際にデータセットから学習されたニューラルネットに対し，提案法を適用することにより，ニューラルネットの大局構造を可視化し，各コミュニティが推論において果たす役割について考察を行う．本実験では，入力画像と，その中に写っている図形の種類 (クラス) が組になった図形認識のためのデータセットを生成し，ニューラルネットの学習に用いた．このデータセットは 10 クラスの図形が描かれた $15 \times 15$ 画素の画像からなり，各クラス 1000 個ずつの学習データを含む．各画像は，それぞれ決められた平均と分散を持つ分布からランダムに生成された点を直線で結ぶことにより二値画像を生成した後，全ての画素に対し平均 0, 標準偏差 0.1 の正規分布から生成した雑音を加えることにより生成した．ここで，各図形において生成する点の x 座標，y 座標における平均値と，直線で結ぶ点の組み合わせを以下のように設定した．ただし，各画像における x 座標, y 座標の最大値を 1, 最小値を 0 とおいた．

- 長方形 (Rectangle)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.2, 0.8)$, 点 c: $(0.8, 0.8)$, 点 d: $(0.8, 0.2)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(b, c)$, $(c, d)$, $(d, a)$ とした．
- ハート型 (Heart): $\cdots$ 点 a: $(0.1, 0.5)$, 点 b: $(0.3, 0.8)$, 点 c: $(0.5, 0.6)$, 点 d: $(0.7, 0.8)$, 点 e: $(0.9, 0.5)$, 点 f: $(0.5, 0.2)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(b, c)$, $(c, d)$, $(d, e)$ $(e, f)$, $(f, a)$ とした．
- 逆三角形 (Triangle)$\cdots$ 点 a: $(0.5, 0.2)$, 点 b: $(0.8, 0.8)$, 点 c: $(0.2, 0.8)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(b, c)$, $(c, a)$ とした．
- バツ型 (Cross)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.8, 0.8)$, 点 c: $(0.2, 0.8)$, 点 d: $(0.8, 0.2)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(c, d)$ とした．
- 斜め線 (Line)$\cdots$ 点 a: $(0.2, 0.8)$, 点 b: $(0.8, 0.2)$ とし，直線で結ぶ点の組み合わせを $(a, b)$ とした．
- ダイヤ型 (Diamond)$\cdots$ 点 a: $(0.5, 0.9)$, 点 b: $(0.9, 0.5)$, 点 c: $(0.5, 0.1)$, 点 d: $(0.1, 0.5)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(b, c)$, $(c, d)$, $(d, a)$ とした．
- 左向き矢印 (Arrow)$\cdots$ 点 a: $(0.4, 0.9)$, 点 b: $(0.1, 0.5)$, 点 c: $(0.4, 0.1)$, 点 d: $(0.9, 0.5)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(b, c)$, $(b, d)$ とした．
- リボン型 (Ribbon)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.8, 0.8)$, 点 c: $(0.8, 0.2)$, 点 d: $(0.2, 0.8)$) とし，直線で結ぶ点の組み合わせを $(a, b)$, $(b, c)$, $(c, d)$, $(d, a)$ とした．
- 顔型 (Face)$\cdots$ 点 a: $(0.3, 0.8)$, 点 b: $(0.3, 0.6)$, 点 c: $(0.7, 0.8)$, 点 d: $(0.7, 0.6)$, 点 e: $(0.2, 0.3)$, 点 f: $(0.8, 0.3)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(c, d)$, $(e, f)$ とした．
- 2 本の水平線 (Two lines)$\cdots$ 点 a: $(0.2, 0.2)$, 点 b: $(0.8, 0.2)$, 点 c: $(0.2, 0.8)$, 点 d: $(0.8, 0.8)$ とし，直線で結ぶ点の組み合わせを $(a, b)$, $(c, d)$ とした．

全ての図形について，上記の平均値を持ち標準偏差が 0.07 の正規分布から各点の座標を生成した．図 1 に学習データにおける各画像の例を，また図 3 に学習データにおける各画像の平均値を示した．

上記のデータセットを用いて多層ニューラルネットを学習し，コミュニティ抽出を行った結果を図 4 に示した．また，コミュニティ間に存在する複数の結合の重みから，結合束の定義を行うことにより，大局構造を可視化した結果を図 5 に示した．また，入力層における各コミュニティに属する画素を図 6 に示した．

ここで，入力データは最大値が 1, 最小値が $-1$, 出力データは最大値が 0.99, 最小値が 0.01 になるように正規化して用いた．また，ニューラルネットの学習における反復数を各学習データごとに 100 回ずつとし，$t$ 回目の反復での学習率 $\eta(t)$ を $\eta(t) = 0.7 \times 10^6/(10^6 + 5 \times t)$ とし，LASSO のハイパーパラメータ $\lambda$ を $5 \times 10^{-7}$ とした．コミュニティ抽出においては，ニューラルネットの結合重みの閾値 $\xi$ を 0.3 とし，各層のコミュニティ数を 10 個とし，反復数を 1000 反復の EM アルゴリズムを 300 回繰り返し，最終反復における対数尤度の値が最大となる回の結果を用いた．

図 2 に，学習の各反復における図形の認識率の変化を示した．

## 5. 考察

図 5，図 6 より，図形認識を行うニューラルネットの構造について，以下のような考察が得られる．

最大重みの結合束を出力層から入力層に向かって辿ることにより，リボン型, 2 本の水平線は，主に Com 4 に含まれる画素の情報を用いて認識されるものと推察される．Com 4 に含まれる画素の部分を見ると，リボン型は白であり，2 本の水平線は黒であることが特徴となっている．また，長方形, ハート型, 逆三角形, 左向き矢印, バツ型, ダイヤ型, 顔型, 斜め線は，主に Com 2 に含まれる画素の情報を用いて認識されるものと推察される．この中でも，長方形，ハート型，逆三角形，左向き矢印，バツ型は，出力層の 1 つ手前の隠れ層における同じコミュニティ (図 6 の右から 4 番目のコミュニティ) に対して最大重みの結合束でつながっているため，このコミュニティに含まれるユニットが，これらの図形の識別に用いられていると考えられる．さらに，これらの図形と斜め線は，出力層の 2

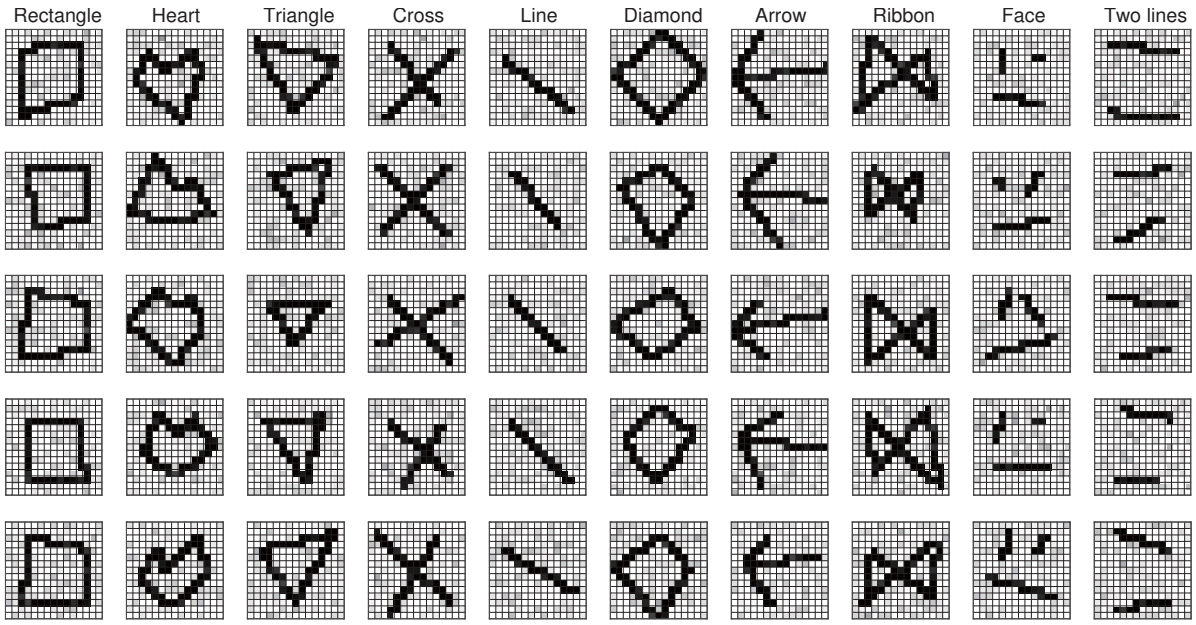


図 1: 学習データとして用いた図形画像の例.

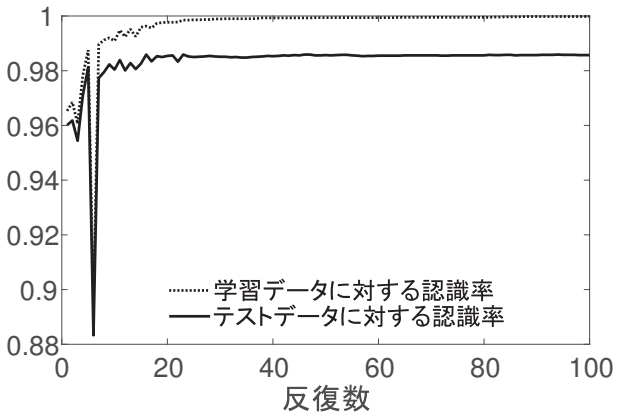


図 2: ニューラルネットの学習における認識率の変化.

つ手前の隠れ層における同じコミュニティ(図 6 の一番右のコミュニティ) に対して最大重みの結合束でつながっており，このコミュニティに含まれるユニットが，これらの図形と斜め線とを識別しているものと考えられる．入力層の Com 9 は入力画像の端の部分にあたり，隣接層への結合束を持たないことから，図形を認識する際には比較的重視されていないことが読み取れる．

## 6. 結論

本研究では，データから学習されたニューラルネットに対し，新たな大局構造の抽出法を適用することにより，各層におけるユニットのコミュニティが入力側のどのコミュニティと強く結びついているかを知ることを可能にした．また，実際にデータセットを用いて実験を行い，ニューラルネットにおける各部分が図形認識において果たす役割の考察を行った．ニューラルネットから抽出された各コミュニティが果たす役割について，更なる詳細な理解を可能とすることや，その結果をニューラルネットの構成の改善に応用することは，今後の課題である．

## 参考文献

[渡邊 17a] 渡邊千紘，平松薫，柏野邦夫．多層ニューラルネットにおける正負の結合重みに基づく大局構造抽出．情報科学技術フォーラム (FIT2017), 2017 年.

[Watanabe 18] Watanabe. C and Hiramatsu. K and Kashino. K. Modular Representation of Layered Neural Networks. Neural Networks. Vol. 97. 2018. pp. 62–73.

[Watanabe 17b] Watanabe. C and Hiramatsu. K and Kashino. K. Recursive Extraction of Modular Structure from Layered Neural Networks Using Variational Bayes Method. Proceedings of Discovery Science 2017, Lecture Notes in Computer Science. Vol. 10558. 2017. pp. 207–222.

[Watanabe 17c] Watanabe. C and Hiramatsu. K and Kashino. K. Modular Representation of Autoencoder Networks. Proceedings of 2017 IEEE Symposium on Deep Learning, 2017 IEEE Symposium Series on Computational Intelligence. 2017.

[Ishikawa 90] Ishikawa. M. A Structural Connectionist Learning Algorithm with Forgetting. Journal of Japanese Society for Artificial Intelligence. Vol. 5. 1990. pp. 595–603.

[Tibshirani 94] Tibshirani. R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B. Vol. 58. No. 1. 1994. pp. 267–288.

[Newman 07] Newman. M and Leicht. E. Mixture models and exploratory analysis in networks. Proceedings of the National Academy of Sciences. Vol. 104. No. 23. 2007. pp. 9564–9569.

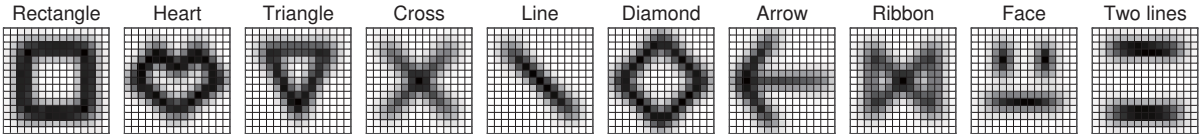


図 3: 学習データとして用いた図形画像の平均値.

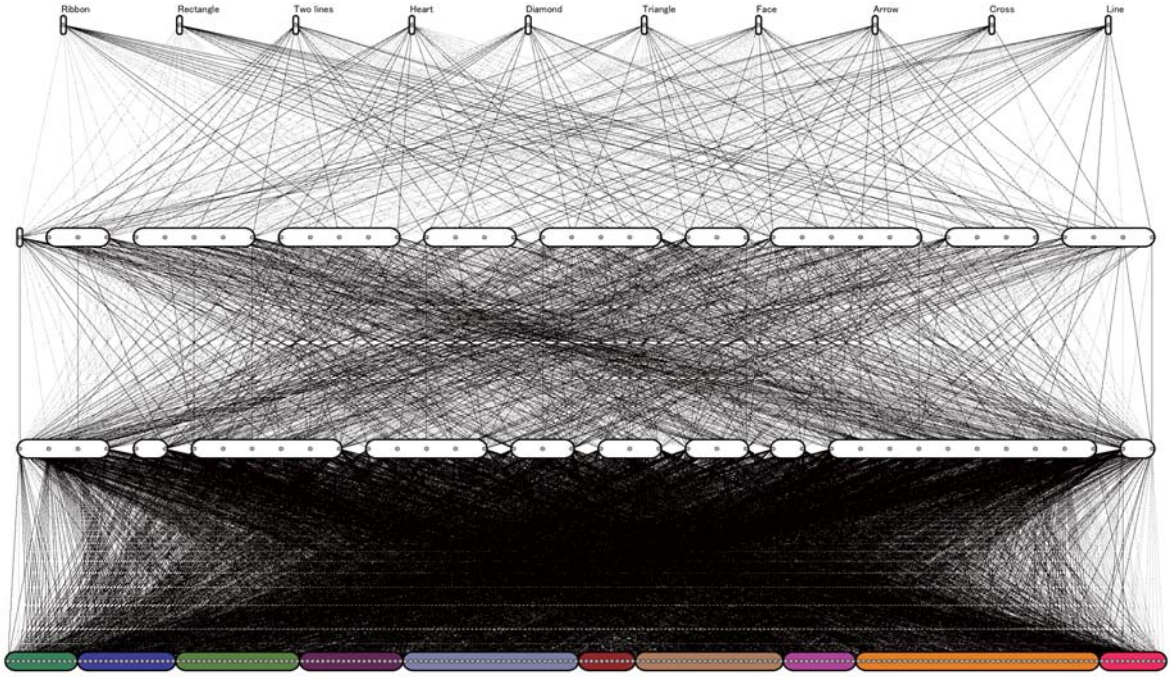


図 4: ニューラルネットから抽出されたコミュニティ構造．入力層の各ユニットは，入力データの画像における 1 画素に対応する.

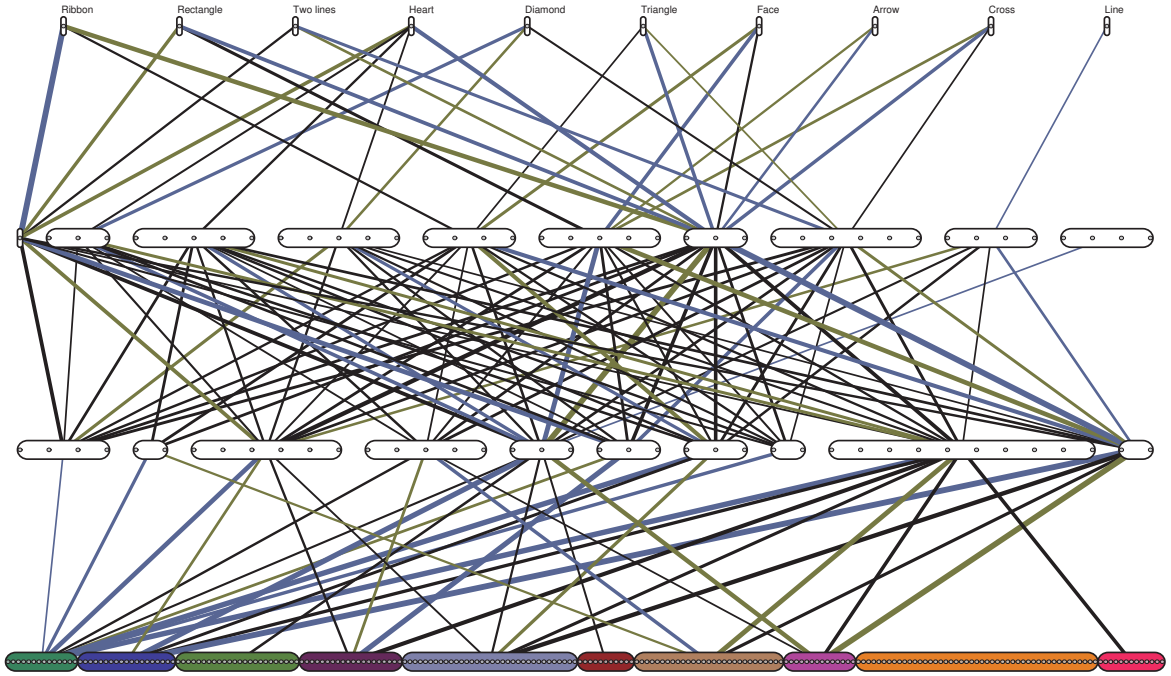


図 5: 各コミュニティ間に存在する複数の結合をまとめて 1 本の結合束として表現した大局構造.

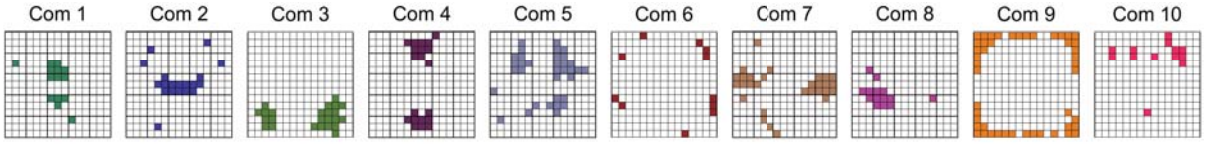


図 6: 入力層の各コミュニティに含まれるユニットと，入力データの画像における画素の対応関係.