# クラウド画像解析エンジンにおけるレイテンシを考慮した DCNN の自動分割法

EP13021 猪子弘康　　指導教授：藤吉弘亘

## 1. はじめに

ロボットの処理をクラウド上のコンピュータで行うクラウドロボティクス [1] への関心が高まっている．これまでに，Deep Convolutional Neural Network(DCNN) によるクラウドロボティクスのための顔画像認識エンジン [2] が提案されているが，DCNN を分割する層を事前に決定する必要があり，状況に応じた負荷分散を効率的に行うことができなかった．本研究ではクラウド型顔画像解析エンジンにおけるレイテンシを考慮した DCNN の自動分割法を提案する．

## 2. レイテンシを考慮した DCNN の自動分割

自動で分割層を決定するには，クライアント，またはサーバの計算不可を低減する，通信量を低減する，など何らかの条件を設定する必要がある．

しかしながら，これらの単独の条件では分割層が一意に決まらない場合がある．そこで本研究では，これらの条件に加え，要求レイテンシを導入する．要求レイテンシは，画像が入力されてから，解析結果を受け取るまで待つことができる許容時間のことである．クラウド型顔画像解析エンジンにおけるレイテンシを考慮した DCNN の自動分割法について以下で詳細を述べる．

### 2.1. システムの構成

本システムの構成を図 1 に示す．ロボット側では入力画像から検出した顔領域を DCNN に入力し，中間層の特徴マップ，ユーザが設定した要求レイテンシと共にクラウドサーバに送信する．クラウドサーバ側では，受け取った特徴マップを DCNN の下位層に入力し，顔画像解析結果を出力する．また，要求レイテンシを元に分割層を決定し，顔画像解析結果と次回選択する分割層の情報をロボットに送信する．

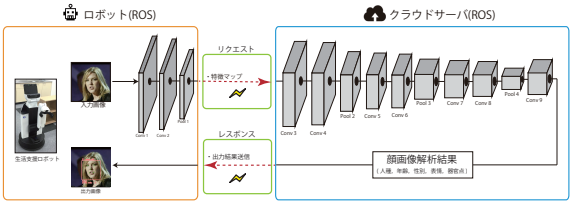



図 1：クラウドによる顔画像の属性推定の流れ

### 2.2. 分割層決定の方針

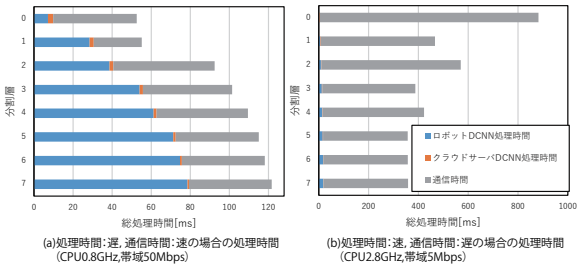
DCNN の分割層を自動的に決定する方法は幾つか考えられるが，我々はクライアントの計算時間 $t_c$，サーバの計算時間 $t_s$，通信時間 $t_t$ の他に要求レイテンシ $R$ の合計 4 つの要素から，DCNN を自動的に分割する．これにより，ユーザのリクエストに可能な限り応えられるシステムとなる．ただし，サーバの負担を小さくするために，要求レイテンシを満たしている場合には，サーバの計算量を小さくするように DCNN の分割層を決定する．これらの条件を式 (1) に示す．この式を用いて次の時刻の分割層 $I$ を決定する．

$$I = \arg\min_i \ t_s(i) \qquad (1)$$

$$s.t. \quad (t_c(i) + t_s(i) + t_t(i)) < R$$

## 3. 予備実験

通信帯域やロボットの性能によって，総処理時間がどのように変化するか調査する．図 2(a) はクライアントの処理性能が低いが，通信帯域が広い場合，図 2(b) はクライアントの処理性能が高いが，通信帯域が狭い場合の総処理時間の関係を示したものである．まず，図 2(a) および図 2(b) より通信帯域が狭いほど通信時間は長くなり，クライアントの処理性能が高いほど処理時間は短くなることがわかる．また，総処理時間は分割層によって差が出ることが確認できる．




(a)処理時間：遅，通信時間：速の場合の処理時間 (CPU0.8GHz,帯域50Mbps)　(b)処理時間：速，通信時間：遅の場合の処理時間 (CPU2.8GHz,帯域5Mbps)

図 2：各条件での処理時間

## 4. 評価実験

提案システムの有効性を検証するために，評価実験を行う．

### 4.1. 実験環境

クライアント，クラウドサーバ共に OS は Ubuntu14.04 LTS で，ROS は Indigo を使用する．クライアントとクラウドサーバはルータを介し 1Gbps で接続する．異なる通信状況下を想定するため，帯域制限を行う．クライアントは，異なる性能の計算機が搭載されたロボットを想定するため，CPU のクロック周波数を変化させる．

### 4.2. 実験結果

要求レイテンシを考慮した分割が理想的にできるかどうか検証する表 1 および表 2 に処理時間及び通信時間を変化させた際の分割結果を示す．

表 1 は，クライアントの処理時間が長く，通信時間が短い場合を示している，これにより，要求レイテンシを満たしつつ，可能な限り出力層側を選択できていることがわかる．

表 1：ロボット性能:低，通信帯域広の場合 (CPU 0.8GHz，50Mbps)

| 分割層 $i$ | 要求レイテンシ [ms] | | | | |
|---|---|---|---|---|---|
| | 60 | 100 | 120 | 300 | 500 |
| 7 | | | | ◎ | ◎ |
| 6 | | | ◎ | ○ | ○ |
| 5 | | | ○ | ○ | ○ |
| 4 | | | ○ | ○ | ○ |
| 3 | | | ○ | ○ | ○ |
| 2 | | ◎ | ○ | ○ | ○ |
| 1 | ◎ | ○ | ○ | ○ | ○ |
| 0 | ○ | ○ | ○ | ○ | ○ |

表 2 は，示すクライアントの処理時間が短く，通信時間が長い場合の結果である．この条件下では要求レイテンシを満たすことができない場合がある．このような場合には，応答時間が最も短くなるような分割層を決定できている．

表 2：ロボットの性能:高，通信帯域：狭の場合 (CPU 2.8GHz，5Mbps)

| 分割層 $i$ | 要求レイテンシ [ms] | | | | |
|---|---|---|---|---|---|
| | 60 | 100 | 120 | 300 | 500 |
| 7 | | | | | ◎ |
| 6 | | | | | ○ |
| 5 | ◎ | ◎ | ◎ | ◎ | ○ |
| 4 | | | | | ○ |
| 3 | | | | | ○ |
| 2 | | | | | |
| 1 | | | | | ○ |
| 0 | | | | | |

## 5. おわりに

本研究では，ユーザのリクエストに可能な限り応え，クラウドサーバの負荷を減らすレイテンシを考慮した自動分割法を提案した．今後の展開として，より最適な分割が可能なモデルの検討などが挙げられる．また，このシステムを基盤とした顔画像解析エンジンの公開を行なった．

## 参考文献

[1] O. Zweigle, *et al*, “Roboearth: connecting robots worldwide”, ICIS ACM, pp. 184-191, 2009.

[2] 山内悠嗣 等,“クラウドロボティクスのための画像認識エンジンの提案”， 電子情報通信学会技術報告，パターン認識・メディア理解，Vol. 115，No. 456，pp. 91-96，2016.