



SEMINAR REPORT

異分野交流が拓く知の創出 —セマンティック通信研究を通じて—



北海道大学大学院
情報科学研究院
メディアネットワーク専攻
准教授

須藤 克弥 氏

本日の講演は、研究者の交流が大事だということを伝えてほしいという依頼を受けて資料を作りましたので、このような大きなタイトルになっています。国内では私がたぶん第一人者だと考えているセマンティック通信の話と、それがなぜ生まれたかという部分をお伝えできればと思っています。

なお、私は博士課程のときに SCAT 研究助成の研究奨励金で毎月 10 万円の支援を頂きました。不思議なことに、ちょうどこのタイミングで私が受け持っている学生が博士課程に行きたいと研究奨励金に応募していて、そういう繋がりもあるのだなと感じています。

改めて、北海道大学大学院情報科学研究院の須藤克弥と申します。基本的には無線通信、情報通信をやっているのですが、そこで AI やコンピュータビジョンを使って研究をしています。

私の実家は岩手の片田舎にありまして、普通にシカがいるような自然の多いところで暮らしていました。

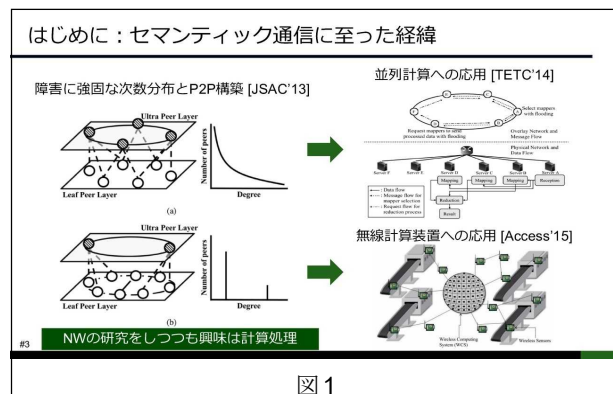
私が中学生のときに携帯電話がはじめて、親にねだって買ってもらいました。ところが、田舎なので電波が全然つながらなかったため、フライパンを携帯の裏に持ってアンテナゲインを強めるとか、そんなことをやっている変な子であったと親にはよく言われていました。

そこからいろいろありまして、東北大学の加藤寧先生のところで博士課程を修了し、その後カナダのウォータールー大学に行き、電気通信大学で助教・准教授をやらせていただき、現在の北海道大学に至ります。職場を変えていく中でいろいろな出会いがあり、研究テーマが少し変わっていったというところも後で紹介いたします。

いろいろと賞もいただいています、特に IEEE VTC 2013 - Spring 最優秀論文賞は先の研究奨励金をもらっていたときに

参加した国際会議で受賞したもので、すごくうれしかったことを覚えています。

図 1 は博士課程のときの研究です。通信の細かい違いは分かりづらいかもしれませんが、ネットワークレイヤーの研究をしていました。その前はグラフ理論をやっていた、そこからネットワークに応用する研究をしていました。それをさらに並列計算へ応用する研究であったり、図 1 右下の無線計算装置への応用にあるボール状のものは中が全部ミリ波になっていて、計算機同士がミリ波で通信しながら計算し合うという謎の装置を作る研究をやってみたり、これをグラフ理論でどのようにリンクを作れば障害性がないかといったような研究をしていました。博士課程まではこうした研究をやっていたのですが、グラフ理論やネットワーク構造というところが得意だったので、その後ウォータールー大学に行きました。



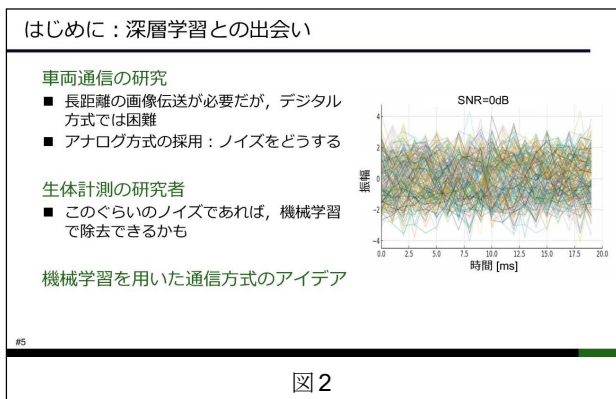
ウォータールー大学のラボのメンバーはトータルで 60 人ぐらい、そのうちポスドクが 30 人ぐらいいたでしょうか。私もポスドクで在籍していました。

このラボのすごいところは、通信がメインなのですが、中でも生体計測、最近ではスマートウォッチで脈から健康度合いを測ることができますが、その最前線の研究をしていたり、車のための通信やスマートグリッドの研究をしていたり、ほかにはごりごりの数学の暗号を研究している人がいたりと本当にいろいろな研究をしている人がいました。そのような研究者と議論する中でセマンティック通信の研究に至ったということになります。

私がウォータールー大学に行ったのは2016年です。そのときは信号機に通信機を設置して、車からの画像をいっぱい集めようというプロジェクトに携わっていました。

ご存じのようにカナダは国土が広く、信号機もすごく密度が低い配置なので長距離の画像伝送が必要になってきます。当時は最大20kmといった長距離が求められていました。そうなるデジタルでの通信方式では難しいのではないかとということで、基本的にはアナログ方式を採用することになりました。

そうして研究を進めていくと、画像を伝送して受信機で復合しようとしたときに、図2のグラフを見て分かるとおりアイバターンが取れないような雑音だらけの信号が送られてくるのです。しかし、私にはこのノイズの取り方の知見がありませんでした。そこでラボのメンバーに相談すると、生体計測の研究者から、生体計測にはすごく雑音が多いらしく、このぐらいのノイズであれば機械学習で除去できるとのアドバイスをもらいました。これが深層学習を使った通信方式を作るきっかけになります。



画像伝送のためのセマンティック通信

ここからは実際に、セマンティック通信とは何かというお話をします。

6Gに向けて車の遠隔監視や工場のロボット操作といった、いろいろなユースケースが増えていく中で、情報を今までのようにアンテナを増やしたり、周波数を高くしたりしてデータレートを上げるだけでは難しく、どうにかしてデータの意味だけを伝える、データを圧縮して送るといった技術が必要になってきました。

図3は、左側の猫と犬の画像をどうやって送るかという例になります。



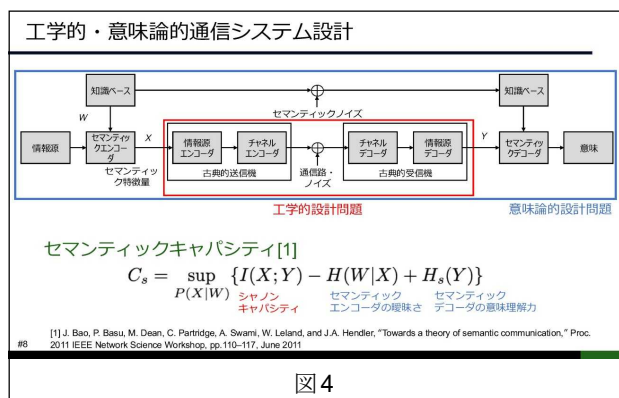
方法Aの古典的な手法では、フーリエ変換で画像を圧縮して、LDPCといった何かしらの通信方式で符号化をして送ることになります。ただし、圧縮にも限界があります。

次に方法Bは「猫が左、犬が右に並んでいる」と画像をテキストに変換します。そして受信者は、その受信したテキストをベースに画像を生成するというものになります。私はこれを「言語による意味の伝送」と定義しています。この方法は、テキストを画像生成AIに入れても、再現性が高くなく、よく分からない画像になってしまうという問題があります。

では、方法Cは何かというと、送信者と受信者が知り合い同士で、同じ知識を持っているのであれば、「タマとポチが並んでいる」という先ほどよりも少ない情報に圧縮することができるというものです。タマとポチの外形を知っていると、タマとポチはいつも左と右に並んでいるということを知っていると、送信者が送った画像のイメージの再現ができるようになります。

そうしたことから同じデータで学習した送信機と受信機があれば、共通認識を基に意味が伝送できるところを最終的なゴールにしています。

ここから少し話が複雑になります。まず、図4の赤枠の部分が70年前にクロード・シャノンが提唱したシャノンキャパシティをベースに設計した工学的な通信方式です。情報をエンコード、すなわちイチゼロに変換し何かしらの符号化方式で送り、そのイチゼロの情報を間違いなく受信機はデコードし復元する。そうして正しい画像が復元できるというものになっています。



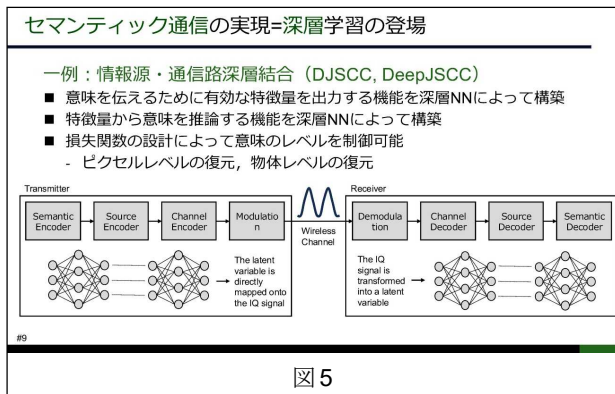
通常はそこまでですが、セマンティックはその概念を少し拡張しています。赤枠の前にセマンティックエンコーダがあって、そこで先ほどのようなテキストの意味や共通認識を持った意味を作るということを行います。このセマンティックエンコーダとデコーダは、基本的に何かしらのデータもしくは知識ベースというものを学習したニューラルネットワークで作られていて、それを基に情報を作るというシステムになっています。

図4下はセマンティックキャパシティの数式で、どれだけセマンティックという概念を使ったら送る情報量が増えるのかを示しています。まず、このI(X;Y)と書いてある、ここが古典的なシャノンキャパシティになります。

このシャノンキャパシティというのは総合情報量なので、基本的には雑音によりどうしてもレートが下がり、キャパシティは減ってしまいます。しかし、このセマンティックのHの部分を導入することで、シャノンキャパシティ以上のキャパシティを達成できるのではないかとということが数学的に示されてい

ます。このような理論的な背景をベースに研究を進めていました。

このセマンティックという概念自体はシャノンキャパシティが出た 70 年前に生まれていましたが、そのときはどうやって意味を作るとか、どうやって定義するということができていませんでした。しかし、深層学習が出てくることによって、機械的にその意味を構築してくれるようなモデルが自動的に作れるようになり実現が可能になったということで、研究開発を進めてきたところになります。



それを一つ実現する方式として、情報源・通信路深層結合 (DJSCC もしくは DeepJSCC) という技術があります (図5)。先ほどの意味を抽出するセマンティックエンコーダという部分と、その意味を圧縮するソースエンコーダという部分と、通信路の符号化をする、誤り訂正をするような部分を全てニューラルネットワークで作ります。それをアナログ変調で送り、受信側は受信したその信号を基に戻し最終的な画像を作るというものになります。これは本当に何も考える必要がなく、何かしらのニューラルネットワークのモデルとデータがあれば、基本的には一番良い通信方式が作れるというものになっています。

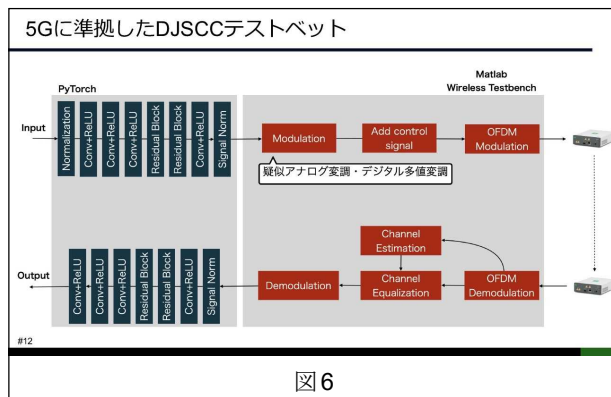
DJSCC の課題 物理層無線方式との親和性

ここからは実際にどのような研究をしてきたかをお話します。

2024 年にソフトバンクが AI-RAN アライアンスを作りました。そこでは基地局に GPU を備えて AI で通信方式を作っていくということで、いろいろな企業が参画しています。

私もそこに参加していて、まだ基地局での実装には至っていないのですが、ソフトウェア無線機という自分でプログラミングし通信方式を作れるものを開発しました。

図 6 が実際に作ったシステム図になります。基本的には移動体通信で使えることを想定しているので、5G に準拠した形になっています。左側に PyTorch を実装していて、先ほど説明したようなセマンティックエンコーダ、ソースエンコーダ、チャンネルコーディングという機能を持ったニューラルネットワークの構造があります。ここから右側の Matlab で変調または OFDM で二次変調して通信するシステムになります。ここが既存の 5G にほぼ準拠した形になります。また、OFDM の復調をした後にチャンネル等化をして、出てきた良い信号に対して復元を試みるというシステムになっています。



ところで、GPU サーバーというと大きいサーバーをイメージされるかもしれませんが、実際、学習のときは大きいサーバーで行いますが、推論のときは図 7 右側のシステム全体像にあるような小さいノートパソコンでも実行できます。それをソフトウェア無線機につないで、実際に通信をして画像伝送することになります。

なお、このままでは電波法に引っかかるので、実験は電波暗室で行っています。暗室はそれほど大きくないので距離は稼げないのですが、減衰器を入れて種々の実験をしています。

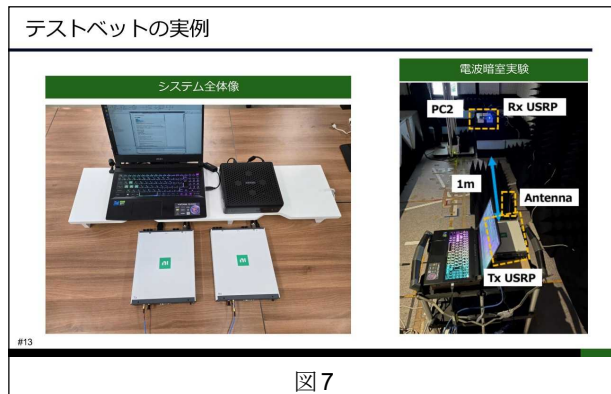


図 8 は実験の結果です。まずは図 8 左側の「SNR が画質に与える影響」から紹介します。ちなみに、圧縮率は 20%です。左上の大きく雑音のがる場合 (SNR=0dB) と、右下の雑音のがらない場合 (SNR=30dB) で比較すると、本来であれば SNR=0dB の場合はなかなか画像は送れませんが、この場合は、バイクと乗っている人物といった情報はしっかり送れています。もちろん、細かく見ると雑音みたいなものが載ってはいますが、画像の持っている意味というか、大事な部分についてはきちんと送れていることが分かります。このことから SNR に関係なく送れるということが強みといえます。

図 8 右側の「遠隔映像伝送」は車のダッシュボードにカメラをつけて、そこから画像を送る実験を行いました。本当はこれを遠隔操作などに使いたいのですが、現時点ではそこまで実現できてなくて画像を送るだけのものになります。

まだフレームレートが悪く、また、SNR=5dB とかなり悪い状況なのですが、割と人が識別できるくらいに見えているので、遠隔監視に使えるのではないかと考えています。なお、無駄な情報は送らないようにしているので、例えば空の色味や雲の大きさといった部分は、元の画像から比べると若干質が悪くなっています。

画像・映像伝送結果

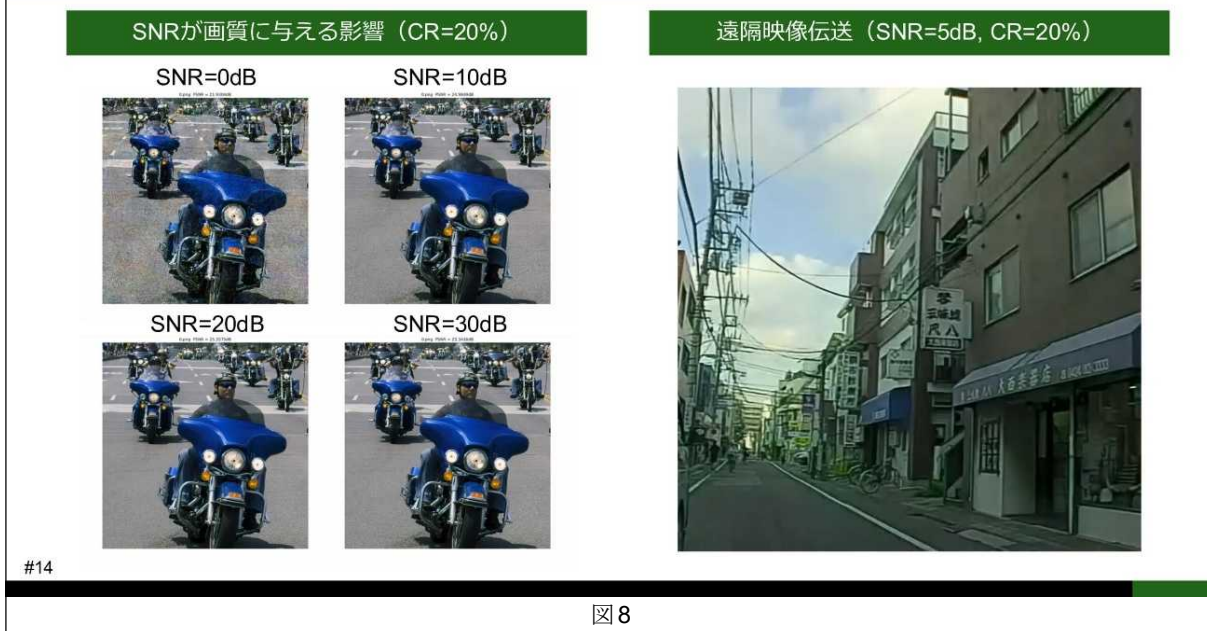


図8

これをさらに遠隔監視につなげるためには、自分が見なければいけない部分を物体検知しバウンディングボックス技術で対象範囲を囲んで表示するのが大事になってきます。

図9はその結果で、左側の「ロスレス画像」は全く圧縮していない画像での物体検知、右側の「DJSCC」は我々が提案しているセマンティック通信方式で送った結果です。両方ともYOLOというAIを使っています。

細かい特性の違いはあるのですが、ほぼ人、車、バイクや自転車は検知できています。このときのYOLOは精度がそもそも悪く85%程度でした。ロスレス画像でも85%だったのですが、DJSCCを適用した場合でも84.9%と0.1%しか下がらなかったため、十分に使える精度であると考えています。

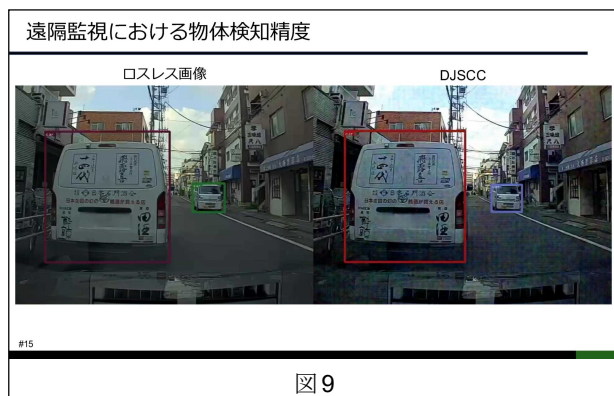


図9

話は戻るのですが、実験の起点は、通信距離を伸ばしたかったことにあります。詳細は省きますが、同等の画質を達成しようとした場合、既存のデジタル変調で送る方法に比べ、疑似アナログ変調で送る方は20倍程度の通信距離を伸ばせるということが図7の電波暗室での実験で分かりました(図10)。これは優れた結果なので、様々なケースでの適応を考えています。

具体的には、最近では高速道路において物が落ちたときの検知用カメラをつけているようで、そういったところの通信方式で使えないかと考えています。

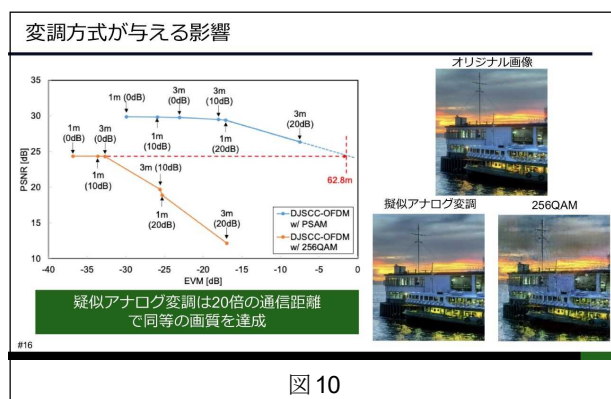


図10

DJSCCの課題 高速移動体への適応

次に、これを高速移動体に適応したいと考えました。具体的にはドローンで撮影した画像を、遠いところへ送るというものです。私は東北大学出身で、入学したのが2011年東日本大震災があったときです。大学では災害に強いネットワークを作る研究もやっていましたので、そういうところに活用できないかと考えました。

細かい部分の説明は省きますが、UAV映像伝送実験例をご説明します。

この実験は山間部の高速道路で行いました。SNR=0dBで本当に信号品質が悪く、距離的にも遠いところになります。ドローンが60km/hで飛んでいるので、フェーディングやドップラーが発生して、信号として扱うのは困難な条件になります。

我々の作った方式と、比較として LDPC+VP8 のかなり圧縮ができる方式で実験を行いました。先ず LDPC+VP8 の方式では動画はほとんど止まってしまいます。最新の画像圧縮技術を使っても、通信路のデータレートに合わないので、ある程度バッファリングしてから部分的なキーフレームだけ送るという形になっています。

一方で、我々の作った方式は、画質は悪いものの、車がどこにいるといった情報・意味は伝えられていることが分かりました。

DJSCC の課題 セキュリティへの対応

残る DJSCC の課題はセキュリティへの対応です。変調をアナログでやっているのだから、基本的に今までの暗号化方式を直接的に使えないところが問題としてあります。そのため、受信者には復元できて、攻撃者に対しては意味が伝わらないような視覚的セキュリティを担保する手法を設計しました。

繰り返しになりますが、基本的に今までの暗号 (AEAD や DSA) は使えないところが原点にあります。ビット系列のイチゼロで処理しているのではなく、アナログ信号のまま変調しているのだから、そのような課題が出てきます。

問題解決にあたっては Encryption then Compression (以下、EtC) と Compression then Encryption (以下、CtE) という2つの手法が挙げられます (図 11)。一つ目の EtC は既存研究で考えられていたもので、画像を暗号化した後、圧縮する方法で送るというものになります。しかしながら、この方法で行くと、図 11 右下のようにほぼノイズのような画像となり機械学習の役には立たず、伝送自体も成立しないというものになってしまいます。

そのため、我々は二つ目の CtE という手法で考えています。まず画像を圧縮し出てきた潜在変数に対して暗号化するというものになります。

DJSCCのセキュリティ

DJSCCは従来の暗号化技術を利用できない

- ビット系列ではなくアナログ信号のまま変調するため
- 画像暗号化技術の利用が必要

Encryption then Compression (EtC)

- 画像を暗号化した後圧縮する方法
- DJSCCでは、暗号化した画像を上手く圧縮できない

Compression then Encryption (CtE)

- 圧縮した潜在変数に対して暗号化する方法
- 暗号化が可逆であれば、DJSCCの再学習が不要

#23 @AINET LAB, UEC

画像暗号化の例




図 11

ところで、そもそも攻撃者がどういうものかを考えておかなければいけません。我々は深層学習で通信を行っています。そうした場合、攻撃者も深層学習ベースで攻撃してくる可能性があるのだから、深層既知平文攻撃モデルというものを考えました (図 12)。

この深層既知平文攻撃モデルとは何かというと、送りたいデータ (平文) と受信したデータ (暗号文) のデータセットがあれば、それをベースに、先ほどの画像圧縮で作った謎の画像でも元の画像に復元できるのではないかとというもので、実際に攻撃モデルとして成り立っています。したがって、この攻撃モデルを持っているアタッカーがいたとしても、正当なユーザーに対しては復元できて、アタッカーに対しては復元できなくなるような手法を取り入れる必要があります。

そこで採用したのが、カオスマップ暗号化というものです (図 13)。これは共通鍵を基本的なものとして、そこからカオス系列を作り、そのカオス系列を元に潜在変数の順番をシャッフルするというものです。そのため、基本的には共通鍵を持っている人にしかきれいに復元できないというものになります。

また、このシャッフルするという工程は計算時間がかかるのだから、その部分を機械学習や深層学習で行うとさらに高速で実現することができます。

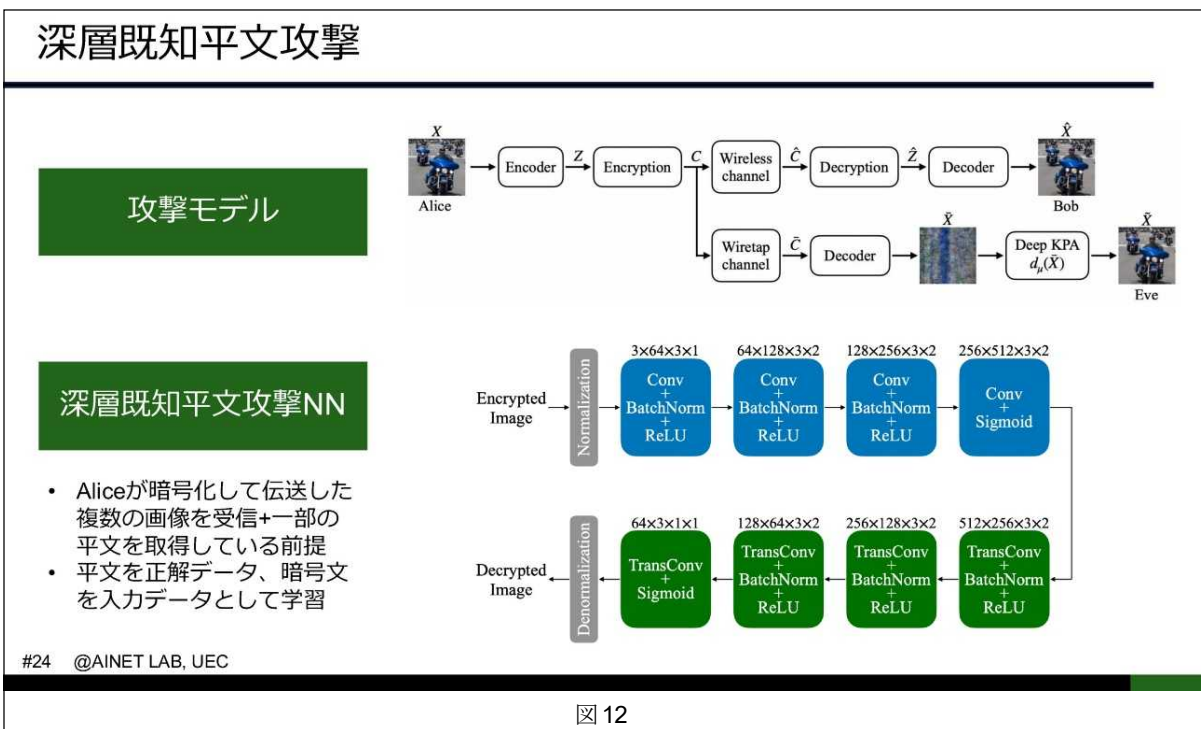


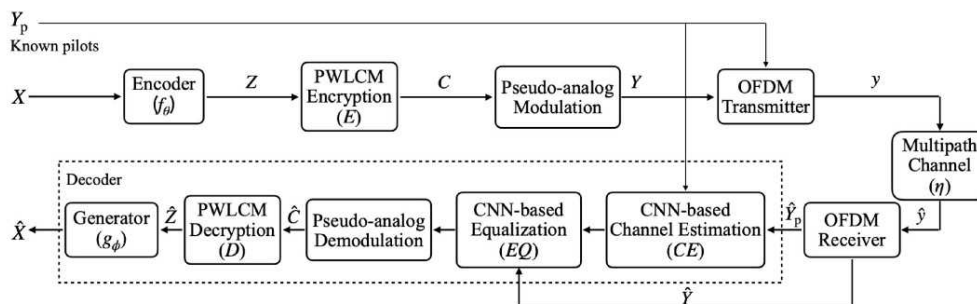
図 12

カオスマップ暗号化を用いたDJSCC

視覚的にセキュアなDJSCC

PWLCM (Piecewise Linear Chaotic Map) を暗号化に利用

- 共通鍵から作成したカオス系列を元に潜在変数を列毎, 行毎にシャッフル
- 鍵が1bit異なるだけで全く異なるカオス系列を作成できるため安全
- 他のカオスマップ暗号よりも高速な処理が可能

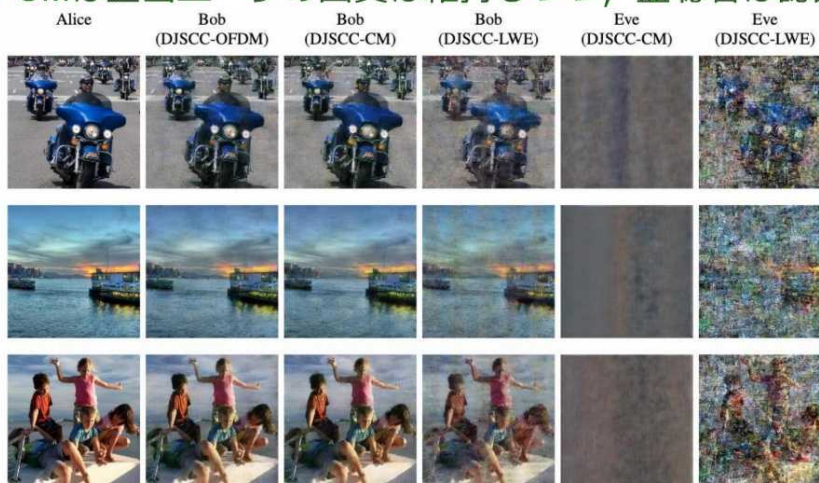


#25 @AINET LAB, UEC

図 13

性能評価

DJSCC-CMは正当ユーザの画質は維持しつつ, 盗聴者は認識が困難



#26 @AINET LAB, UEC

図 14

そして、この性能評価をしたものが図 14 になります。左側の Alice が送信者になります。Bob は正当なユーザーで、この人には確実に画像を送りたい、右側の Eve はアタッカーで、この人には画像を見せたくないとします。

我々の提案は DJSCC-CM というもので、正当ユーザーに対しては元々の画像と似たものを送り、復元できる一方で (Bob(DJSCC-CM))、アタッカーに対しては全く復元できないものを送る (Eve(DJSCC-CM)) 手法になります。

なお、既存手法は DJSCC-LWE という、画像に少しノイズを加えながら学習するものでした。この場合、正当ユーザーも画質はよくはならず (Bob(DJSCC-LWE))、かつ、アタッカーに対しても、ある程度は見えなくなるものの、ぼんやりとは物や人の形が分かるので (Eve(DJSCC-LWE))、意味としては伝わる可能性を残してしまいます。

このことから、DJSCC-CM は“意味”を伝えないことに成功した初めての技術と言えます。ここまでの 2017 年から 2024 年までのセマンティック通信の研究成果になります。

AI 連携のためのセマンティック通信

最近AIがどこにでも使われるようになってきました。セマンティック通信を活用してAI同士でデータをやりとりする、延いてはより優れたAIを自動的に作っていくところを最後にお話しします。

先ず前者についてですが、例えば工場では、製造ライン上のロボットアームの制御や工場内を移動するロボットの自動運転制御機構などについているセンサーを始めとした、いろいろな部分でAIを搭載しています。そのようなリアル空間における工場を、サイバー空間上に再現し、工場に存在するデータ群の意味を定義していき、全空間的にどのようなAIがあれば工場が一番上手く有機的に機能するか、というモデルを作っていきます。これは、一昨年ぐらいから広まった空間知能にも通ずる概念だと思いますが、その中にセマンティック通信を使っていこうという取り組みになります(図15)。

なお、データの伝送の部分はセマンティック通信により種々の意味を送り、その中で大事な意味は何かというのを蓄積していけばよいということです。

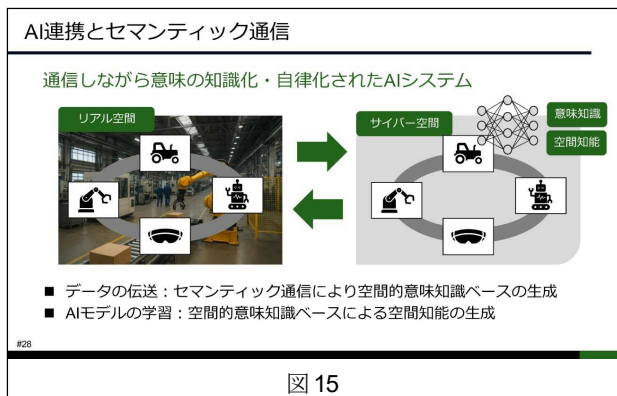


図15

では、もう一つのAIモデルの学習について紹介します。

工場などに多くの端末がある中で、AIモデルの学習に使われているのが、分散連合学習というものになります(図16)。これは何かというと、多くの端末が学習データをやりとりするのは、データ量が増えたり、セキュリティの問題があるので、各端末にAIのモデルを送りローカルで学習した後、各端末の結果を集約して学習していこうという手法になります。

ただし、近年はどうしてもこのAIモデル自体が線形にデータ量が増加していく中で、今は学習データとAIモデルのどちらのやり取りが良いのか分からない状況にあるものの、我々は、このAIモデルを送るという部分について何とか上手く圧縮するであったり、モデル自体の意味を抽出したいと考えました。

分散連合学習への適用

分散機械学習は低コストで良いAIモデルを構築するための基盤技術

- ノード間で学習データではなく、モデルを共有
- しかしながら、近年はモデルサイズが増加傾向(2012年から10倍)

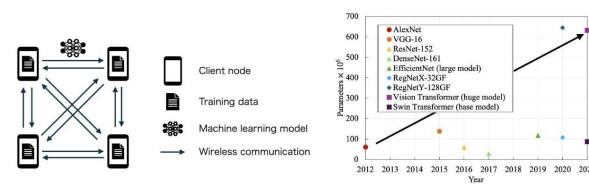
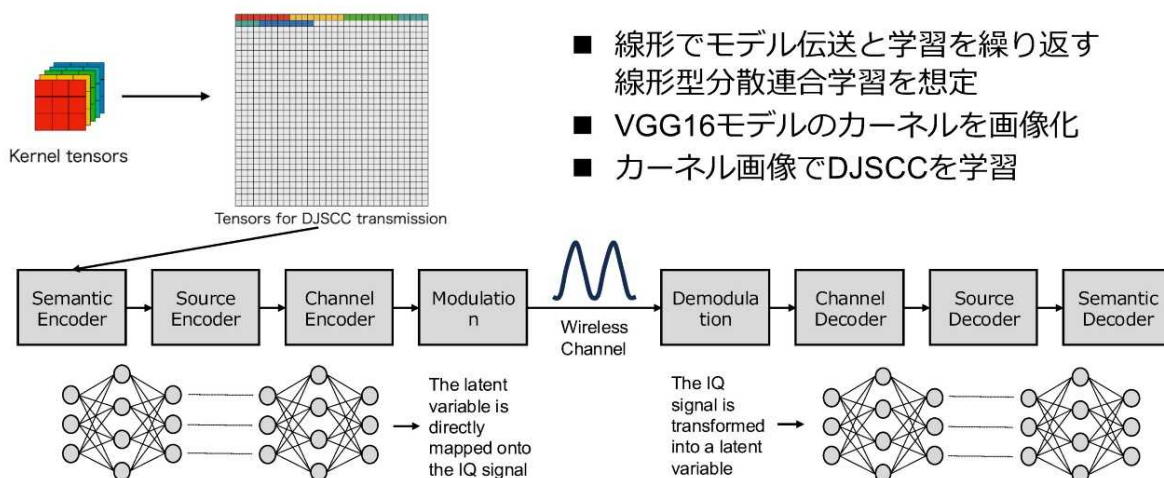


図16

行ったことは簡単で、先ほど挙げたDJSCCを適用したものになります。例えば、VGG16というニューラルネットワークがあるのですが、そこで使われている畳み込みのカーネルを画像イメージになるように埋め込んでいくだけのデータを作り、それで学習してから通信機を構成して、実際に分散連合学習のところで適用しました。(図17)

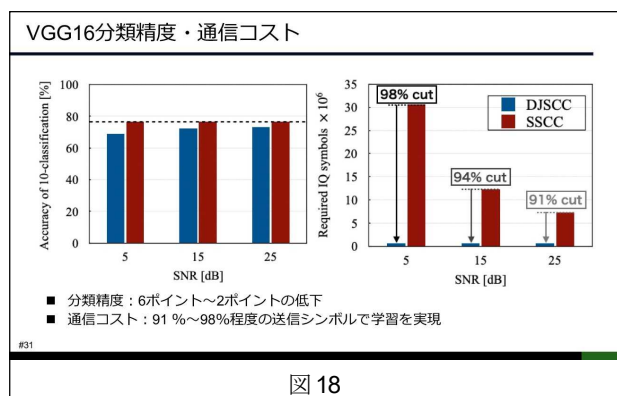
DJSCC適用型分散連合学習



#30

図17

図18がそのときの分類精度と通信コストで、DJSCCの結果は青色で示しています。左の分類精度グラフでは普通にデジタルで送るより精度が少し下がることを示しています。必要なデータは送っているものの、やはり意味を送るというだけでは、分類精度は少しだけ性能が下がってしまいます。ただし、送らないといけない信号数は最大で98%程度カットできているので、通信のコスト面はかなり削減できることが分かりました。この結果により、今までご紹介した手法とセマンティック通信を複合する形で研究を進めているところになります。



今日は、セマンティック通信の背景や、なぜ自分がこれに行き着いたかという部分を紹介しました。

生成AIは世界モデルと言うデータ群に対する汎化性を上げていくのに対して、我々が取り組んでいるセマンティック通信は、使いたい場所や通信品質の条件に合わせて作る生成AIだと思ってもらえれば大丈夫です。

今後の課題としては、送信者が意味をデータで送るとして、受信者はその送られたデータに意味があるかどうかをどのように判断するのか。つまり、従来の通信であれば、ビットの一個でも誤りがあれば再送をしなければいけなかったのですが、そういうところをどうやってプロトコルとして今後標準化していくかが課題としてあります。そして今はJSTの先端国際共同研究推進事業(ASPIRE)で、オウル大学の先生と協力して標準化に持っていけないかと考えています。

私のやっている研究は、アプリケーションに合わせて通信方式が設計できるようになっています。そのため協調型ロボットや自動運転といった、様々なニーズに合わせて実用化できればと考えています(図19)。

発表は以上になります。ありがとうございました。

おわりに

セマンティック通信とは

- 情報が持つ意味を見つけることであり、生成AIと同じ思想
 - 生成AIが世界モデルに対する汎化性
 - セマンティック通信は局所・局時的な条件の考慮も必要
- セマンティック通信における再送とは
 - 意味が伝わったかどうかをどのように送信者は認識するのか？
- アプリケーション志向の通信方式
 - 協調型ロボット, 自動運転, etc
 - 他分野での通信の課題・ニーズを議論できればと思います

図19