

修士論文

LHC-ATLAS 実験における
Transformer を用いた
タウ粒子同定手法の開発

早稲田大学
先進理工学研究科 物理学及応用物理学専攻
寄田浩平研究室

吉田大起

2025 年 2 月 7 日

概要

素粒子物理学において標準模型は粒子のふるまいを良く説明できる優れた理論である一方で、実験での検証から理論と観測事実の矛盾が指摘されている。そのため現在も精密測定による理論の検証や、矛盾を説明できる新物理の探索が進められている。

欧州原子核研究機構（CERN）では周長 27km に及ぶ陽子・陽子衝突型加速器、LHC を用いた複数の実験を行っており、そのうちの 하나가 ATLAS 実験である。LHC の陽子衝突点の一角所に ATLAS 検出器を設置し標準模型の精密測定や新粒子探索を目的としている。

これらの解析において重要となる粒子の 하나가 タウ粒子である。タウ粒子を使う精密測定の例としてはヒッグス粒子の精密測定があり、背景事象と分離がしやすく崩壊分岐比も比較的大きいタウ粒子への崩壊チャンネルは有用である。新粒子探索の例としては超対称性理論で言われている重いヒッグス粒子の探索があり、タウ粒子対への崩壊分岐比が大きいという予想があるため今後の解析に必要となる。

タウ粒子は唯一のハドロン崩壊が可能なレプトンであり、かつ寿命が短い粒子である。ATLAS 実験においては検出器到達前に大多数が崩壊し、タウ粒子として再構成できるのはハドロン崩壊したもののみである。しかしハドロンを生じる過程は他にも多く存在するため区別が難しく、ATLAS 実験では機械学習を用いた手法によりタウ粒子同定を行うことで解析感度を向上させてきた。現行手法では深層学習の一種である RNN (Recurrent Neural Network) がタウ粒子の同定に使用されている。

本研究では近年注目されている深層学習モデルの Transformer を、同定手法に取り入れることでタウ粒子の解析精度を向上させられるか検討を行う。Transformer は Attention 機構により特徴量の中から学習に有効なものを強調することで学習能力を高めたモデルである。タウ粒子の同定に使用する特徴量は大きく分けてトラック、クラスター、ジェットの変数である。トラックは検出器で再構成された粒子の飛跡に関する情報である。クラスターは粒子のエネルギーを測定するカロリメータで得られた情報である。ジェットはタウ粒子とその他の QCD ジェットを分類するのに有効な特徴量が含まれている。

モデルの学習と評価にはシミュレーションデータが用いられており $\gamma^* \rightarrow \tau\tau$ を信号事象、dijet を背景事象に使用している。既存の手法と同様にタウ粒子の崩壊の仕方に合わせて Transformer モデルを学習させることにより、現行手法よりも優れた背景事象除去能力を持つことを示した。また、現行手法では RNN で行っているタウ粒子の prong 数の判別を、Transformer で行い改善できるか検証したところ有効であることを示した。

今後の課題としては入力変数とモデルのパフォーマンスの関係性を調べることや、実験データに対しても今回使用したモデルが有効か検証する必要がある。また、prong 数まで判別する Transformer モデルではタウ粒子の電荷を決めることができないため、解決方法を考える必要がある。これらの課題を解決し最終的には精密測定や新物理探索の解析感度向上を目指す。

目次

1 序論	7
1.1 標準模型	7
1.2 標準模型を超えた物理	8
2 LHC-ATLAS 実験	9
2.1 LHC	9
2.2 ATLAS 検出器	10
2.2.1 ATLAS 実験で用いられる座標系	11
2.2.2 内部飛跡検出器	11
2.2.3 カロリメータ	12
2.2.4 ミューオンスペクトロメータ	13
2.3 トリガーシステム	14
3 タウ粒子	15
3.1 性質	15
3.2 タウ粒子を使用した解析	15
3.2.1 ヒッグス自己結合定数の測定	16
3.2.2 Heavy Higgs の探索	17
3.2.3 スタウ探索	18
3.3 LHC-ATLAS 実験におけるタウ粒子の再構成手法	20
3.3.1 タウ粒子再構成	20
3.3.2 トラック識別	20
3.3.3 タウ粒子同定	23
3.4 タウ粒子同定手法の先行研究	26
4 本研究で取り入れるタウ粒子同定手法	28
4.1 Transformer	28
4.2 GNN	30
5 Transformer と現行手法の比較	31
5.1 学習で使ったサンプルと条件	31
5.2 モデルの構造	31
5.3 モデルの学習方法	32
5.4 性能比較	33
6 タウ粒子を崩壊後の prong 数で分けずに学習を行った Transformer モデル	35
6.1 学習と評価方法	35
6.2 性能比較	36
7 prong 数の判別とタウ粒子同定を行う Transformer モデル	38
7.1 使用するモデルと分類方法	38
7.2 背景事象除去能力の新たな定義	40
7.3 背景事象除去能力の比較	40

8 今後の展望	46
8.1 まとめ	46
8.2 考察	46
8.3 今後の展望	47
9 謝辞	48
A Transformer モデルで次元数と層数を変えた場合の性能比較	49
B 横運動量ごとの背景事象除去能力	51
C タウ粒子同定と prong 数の判別を行う Transformer での score 分布	52
D 2prong も利用した RNN+Transformer と Transformer inclusive の比較	53
E 入力変数を絞った場合の性能比較	55
参考文献	57

図目次

1.1	標準模型 [1]	7
2.1	LHC 概略図 [3]	9
2.2	LHC/HL-LHC の実験実績及び実験計画 [4]	10
2.3	ATLAS 検出器概略図 [5]	10
2.4	内部飛跡検出器概略図 [6]	12
2.5	カロリメータ概略図 [5]	13
2.6	ミューオンスペクトロメータ概略図 [5]	13
2.7	LHC-ATLAS 実験 Run2 におけるトリガーシステム [9]	14
3.1	タウ粒子崩壊のファインマンダイアグラム	15
3.2	ヒッグス対生成の起こるファインマンダイアグラムの例 [11]	16
3.3	μ_{HH} の予測値と測定値 [11]	16
3.4	$\tau_{lep}\tau_{had}$ と b-tag を使用した場合の m_T^{tot} 分布 [12]	17
3.5	$\phi(h/H/A) \rightarrow \tau\tau$ の断面積に対する上限 [12]	18
3.6	スタウ探索に利用される崩壊事象 [13]	18
3.7	タウ粒子追加の運動量から作成した m_{T2} 分布 (左: low mass 領域、右: high mass 領域) [13]	19
3.8	スタウと最も軽いニュートラリーノの質量に付けられた制限 [13]	19
3.9	タウ粒子再構成から粒子同定までの手順	20
3.10	トラック識別の精度 (左: truth 1prong tau、右: truth 3prong tau) [14]	23
3.11	truth tau に対し再構成できたタウ粒子由来のトラックの本数 [14]	23
3.12	RNN のタウ粒子同定モデル [18]	24
3.13	ΔR_{max} の変数分布	24
3.14	現行手法でのタウ粒子同定性能 [14]	26
3.15	GNN を用いた先行研究での RNN との性能比較 [21]	27
4.1	Transformer の構造 (左半分が Encoder、右半分が Decoder を表す) [23]	28
4.2	Multi-Head Self-Attention の構造 (右は Scaled Dot-Product Attention の中身)[23]	29
4.3	GraphSAGE の計算手法 [26]	30
4.4	SAGEConv のイメージ図	30
5.1	Transformer を使用したタウ粒子同定モデル	31
5.2	GNN を使用したタウ粒子同定モデル	31
5.3	学習から評価までの流れ	32
5.4	信号事象と背景事象での横運動量分布	33
5.5	各モデルの背景事象除去能力	34
6.1	サンプルを prong 数で分けない場合の Transformer の学習と評価手順	35
6.2	prong 数で分けずに学習させた Transformer モデルと RNN の背景事象除去能力	36
7.1	prong 数の判別とタウ粒子同定を行う Transformer モデルの同定手順	38
7.2	prong 数の判別も行う Transformer	38
7.3	トラック識別とタウ粒子同定を並行して行う GNN モデル	39
7.4	prong 数の判別能力まで含めた背景事象除去能力の比較	40
7.5	各モデルでの prong 数判別結果 (左上:RNN、右上:GNN、下:Transformer)	42
7.6	正解の Tau Track の本数分布	43
7.7	各モデルにおける正解の Tau Track 数に対する prong 数の判定結果 (上段:RNN、中段:GNN、下段:Transformer)	44

A.1	次元数と層数を変えた場合の比較 (左: 次元数を変動、右: 層数を変動)	49
A.2	学習時の epoch 数に対する loss の推移	50
B.1	1prong 事象でのタウ粒子横運動量に対する背景事象除去能力 (左:Loose、中央:Medium、 右:Tight)	51
B.2	3prong 事象でのタウ粒子横運動量に対する背景事象除去能力 (左:Loose、中央:Medium、 右:Tight)	51
C.1	truth 1prong tau に対する各 score の分布	52
C.2	truth 3prong tau に対する各 score の分布	52
C.3	background に対する各 score の分布	52
D.1	2prong 事象も利用した RNN+Transformer と Transformer inclusive の比較	53
E.1	jet の入力変数を使用しない場合の Transformer モデル	55
E.2	学習に使用する入力変数を一部除外した場合の比較 (左:1prong、右:3prong)	55

表 目 次

3.1	タウ粒子の代表的な崩壊モード	15
3.2	トラック識別に使用する入力変数 [14]	22
3.3	タウ粒子同定で使用している入力変数 [14]	25
3.4	タウ粒子同定の信号取得効率の基準点	26
5.1	タウ粒子同定モデルの学習に使用した事象数	32
5.2	各基準点での背景事象除去能力	34
6.1	prong 数で分けずに学習させた Transformer における各基準点での背景事象除去能力	37
7.1	各基準点での背景事象除去能力	41
A.1	次元数、層数を変えた場合の各基準点での背景事象除去能力	49
D.1	2prong 事象を利用した場合の背景事象除去能力	54
E.1	各基準点での背景事象除去能力	56

1 序論

1.1 標準模型

素粒子物理学は宇宙を構成する最小単位である素粒子と、粒子間に働く相互作用を定式化することが目的の学問である。これを良く満たすモデルとして標準模型が挙げられる。標準模型は 17 種類の素粒子で構成されており、これらの素粒子は実験によって存在が確認された。それを図 1.1 に示す。

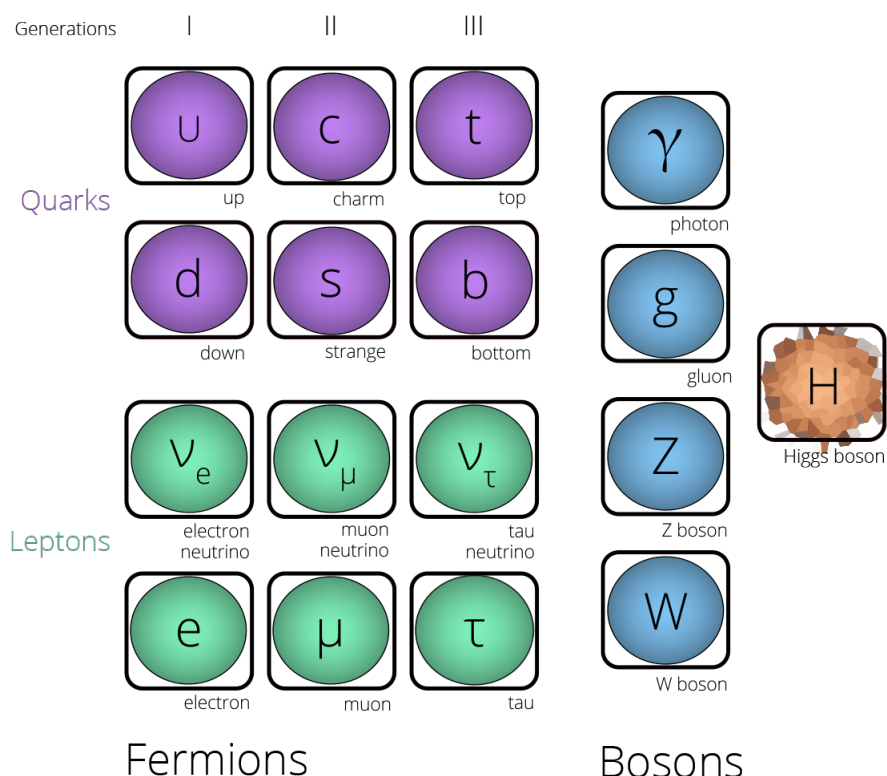


図 1.1: 標準模型 [1]

構成要素はフェルミオンとボソンの大きく 2 項目に分けられる。

1. フェルミオン

スピンの半整数値である $1/2$ をとりクォークとレプトンに大別される。電子に対する電子ニュートリノのように対になった粒子がクォーク、レプトンそれぞれで 3 ペアずつ存在する。このペアを世代と呼び 3 世代存在する構造をとっている。フェルミオンに含まれるすべての粒子は質量を持ち、同じ電荷を持つ粒子同士では世代が大きいものほど質量が大きい。

(a) クォーク

電荷が $+2/3e$ であるアップクォーク、チャームクォーク、トップクォークと、電荷が $-1/3e$ であるダウルクォーク、ストレンジクォーク、ボトムクォークから構成される。

電荷が整数値でないため単独では発見されず、複数のクォークが集まったハドロンとして観測される。色荷を持ち強い相互作用を行う。

(b) レプトン

電荷を持たない電子ニュートリノ、ミューニュートリノ、タウニュートリノと電荷が $-e$ である電子、ミューオン、タウ粒子で構成される。色荷を持たないため強い相互作用を行わない。

2. ボソン

スピンは整数値をとり、相互作用を媒介する。

スピン 0 の粒子として、電磁相互作用を担う光子、強い相互作用を担うグルーオン、弱い相互作用を担う Z ボソン、W ボソンがある。Z ボソンと W ボソンは質量を持ち、前者が 91.2GeV 、後者が 80.3GeV の質量を持つ [2]。

最後に残ったスピン 1 の粒子がヒッグス粒子であり 125.2GeV の質量を持つ [2]。W ボソンや Z ボソンが質量を持つ理由を説明するのに必要なヒッグス機構の導入のために予言された粒子である。ヒッグス粒子は 2012 年に発見され、発見時期も比較的新しいためより精密な測定が求められている。

1.2 標準模型を超えた物理

標準模型が支持される一方で標準模型では説明できない事項がある。ここではいくつか例を紹介する。

● 暗黒物質

宇宙には光の放出、吸収を行わず質量を持った物質が存在することがわかっており、これを暗黒物質と呼んでいる。暗黒物質の存在を示唆する証拠は複数存在する。例として重力レンズ効果がある。これは光が大きな質量を持った天体などのそばを通ると、重力場の影響で進路が曲げられるというものであり、質量を持った物質の分布が確認できる。暗黒物質の候補として weakly interacting massive particles (WIMP) と sterile neutrino がある。

WIMP は弱い相互作用と重力相互作用のみを行い、質量は数 10GeV と予測されている粒子である。直接探索や、WIMP の対消滅後に生じ得る粒子を探す間接探索が行われている。

sterile neutrino は標準模型のニュートリノとは異なり重力相互作用のみを行うと想定されるニュートリノである。質量が軽ければ宇宙に散らばってしまうため、ある程度密集しているという観測事実を満たすには質量が大きくなければならない。重力相互作用しか行わないため崩壊後の X 線を間接探索する必要がある。

● 重力の存在

標準模型では電磁相互作用、強い相互作用、弱い相互作用を扱うことに成功しているが、重力を扱うことができない。この問題に対する説明として試みられているのが力の統一である。これは 4 つの力が宇宙誕生当初は 1 つの力として存在しており、エネルギースケールの低下とともに分岐していった、とする仮説である。現状は重力相互作用を含めた統一の前に、電弱相互作用と強い相互作用を 1 つに統一する大統一理論の検証が課題となっている。

2 LHC-ATLAS 実験

LHC-ATLAS 実験は、欧州合同原子核研究機構 (CERN) で行っている Large Hadron Collider (LHC) を利用した実験の 1 つである。

2.1 LHC

LHC はスイスとフランスの国境付近地下 100m に位置する周長 27km の陽子・陽子衝突型円形加速器である。陽子が密集したバンチと呼ばれる状態を作り出した後、複数のリングを通過しながら電磁場により亜光速まで加速、4 つある衝突点にて 40MHz の頻度でバンチ同士を衝突させる。これらの衝突点では CMS 実験、LHCb 実験、ALICE 実験、そして ATLAS 実験が行われている。ATLAS 実験と CMS 実験はどちらも大まかな構成が近い汎用検出器を使用している。異なる実験での結果を比較することで実験精度を上げることが可能となる。

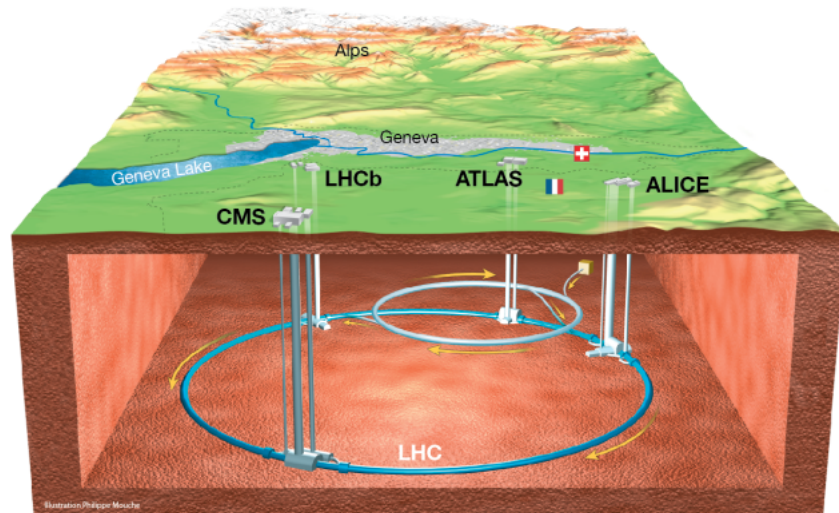


図 2.1: LHC 概略図 [3]

LHC はこれまでに三度の Run を実施しているが、Run と Run の間にある長期休止期間にアップグレードが進められ、陽子衝突の重心系エネルギーは向上され続けてきた。2010 年から 2013 年まで行われた Run1 では $\sqrt{s} = 7 \cdot 8 \text{ TeV}$ 、2015 年から 2018 年の Run2 では $\sqrt{s} = 13 \text{ TeV}$ 、2023 年から続く Run3 では $\sqrt{s} = 13.6 \text{ TeV}$ に到達している。

また、今後は LHC を高輝度化する HL-LHC の建設計画が進められている。これによる衝突頻度の増加により生成断面積の小さい事象の観測が期待される。HL-LHC の計画は様々な要因で遅れが生じたが、現在は 2030 年からの稼働を目指し準備が進められている。LHC と HL-LHC の実験計画を図 2.2 に示す。



図 2.2: LHC/HL-LHC の実験実績及び実験計画 [4]

2.2 ATLAS 検出器

ATLAS 実験では LHC における衝突点の 1 つに設置された ATLAS 検出器で陽子・陽子衝突事象を観測し、標準模型の精密測定や新物理探索を行う。

ATLAS 検出器は円筒形のような外形をしており、高さ 25m、長さ 44m に及ぶ。検出器の構造は大まかに分けると内側から内部飛跡検出器、カロリメータ、ミュオン検出器で構成される。この検出器によって荷電粒子やハドロン、光子等の電荷、運動量、向きなどが測定可能となり、これらの情報を用いて解析を行う。

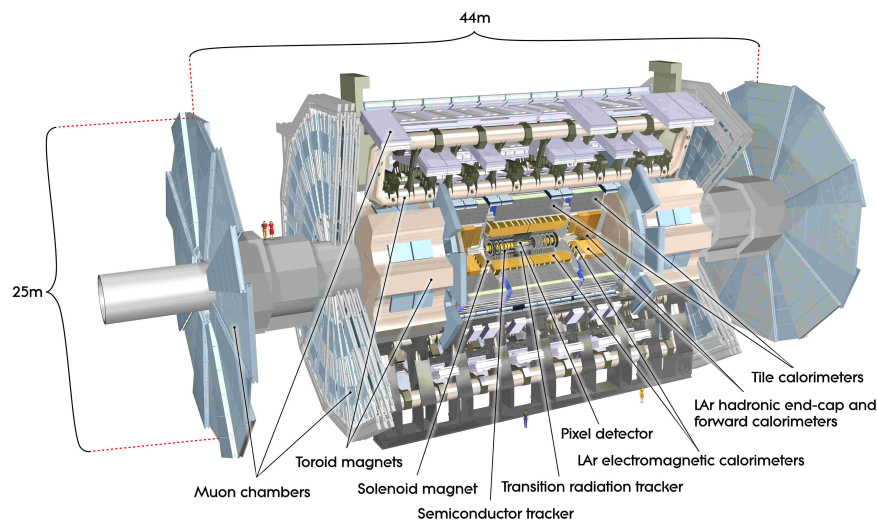


図 2.3: ATLAS 検出器概略図 [5]

以降では ATLAS 実験のかなめとなる各検出器やトリガーシステムについて記述する。

2.2.1 ATLAS 実験で用いられる座標系

ATLAS 検出器の構造について書く前に、ATLAS 実験で用いられる座標系について説明する。ATLAS 実験では陽子ビームの衝突点を原点、ビーム方向に z 軸をとり、円筒座標系 (R, ϕ, z) が用いられる。また、 z の代わりに擬ラピディティ η を用いた記法 (R, ϕ, η) が慣例的に使用される。ここで $\theta \in (0, \pi)$ は z 軸からの仰角であり

$$\eta = -\log\left(\tan\frac{\theta}{2}\right)$$

と変換される。また、粒子の運動量と向きを表現する場合は、 (p_T, ϕ, η) がよく用いられる。ここで直交座標系の運動量成分 (p_x, p_y, p_z) を用いると

$$p_T = \sqrt{p_x^2 + p_y^2}$$

である。他に、座標系とは異なるが粒子間の距離を表す物理量として ΔR があり、2 粒子間の ϕ と η の差、 $\Delta\phi$ 、 $\Delta\eta$ を用いて以下のように計算される。

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

2.2.2 内部飛跡検出器

内部飛跡検出器は $|\eta| < 2.5$ の範囲をカバーし荷電粒子の衝突点や運動量などを測定する。Pixel 検出器、SCT、TRT から構成され、その外側にある超電導磁石によって 2T の磁場がかかっている。曲げられた飛跡の曲率から運動量の測定が可能となる。

- pixel 検出器

半導体によって構成されるセンサーがピクセル上に並ぶことにより、高い粒度を持って測定可能な検出器である。検出器の外形を円柱に例えたとき、側面に当たる部分をバレル、底面に当たる部分をエンドキャップと呼ぶが、それぞれ 3 層ずつで構成される。さらに Run2 からはその内側に the insertable B-Layer (IBL) が挿入されることによってより粒度を上げることに成功した。IBL がビームパイプ中心から $R=33.25\text{m}$ に位置し、その外側に pixel 検出器のバレル部が続く。

- SCT

薄板状のシリコンセンサー 2 枚を重ねたモジュールにより構成される検出器である。センサーにはストリップと呼ばれる帯状の電極が $80\mu\text{m}$ 間隔で並行に張り巡らされており、このセンサーを 40mrad 傾けて 2 枚組にすることにより斜交座標を作ることで通過位置を読み出す。SCT ではバレル部は 4 層、エンドキャップ部は 9 層で構成される。

- TRT

ガスが封入された直径 4mm のストローチューブを束にした検出器である。ガスは Xe、 CO_2 、 O_2 で構成されており、ここを荷電粒子が通過することで生じた電離電子を測定する。また、ストロー間にはポリプロピレン繊維があり、荷電粒子が通過する際に生じた遷移放射の情報が粒子識別に利用される。バレル部は 3 層、エンドキャップは 20 の構成要素からなる。

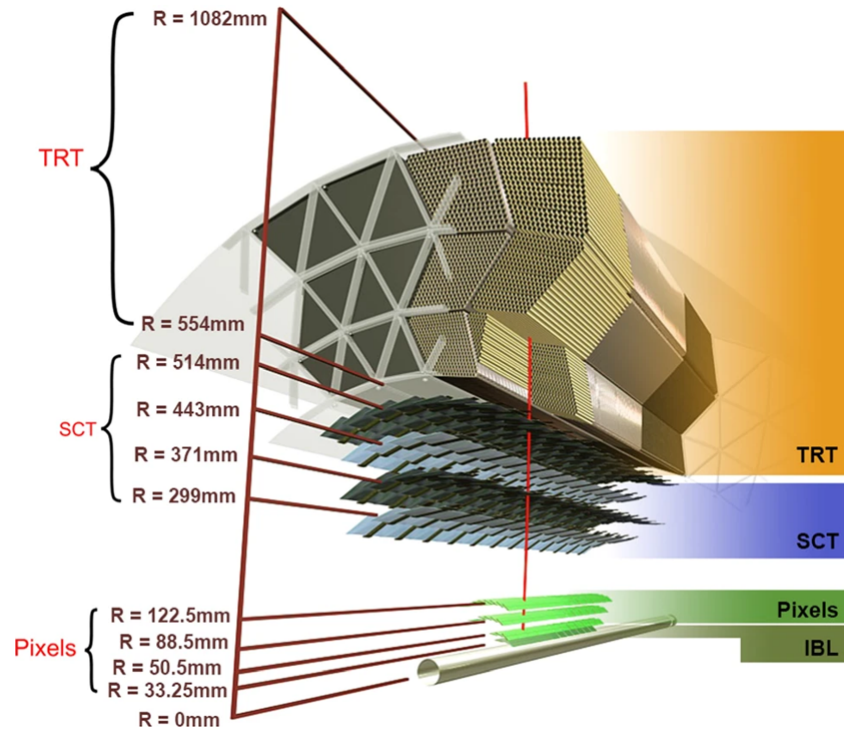


図 2.4: 内部飛跡検出器概略図 [6]

2.2.3 カロリメータ

電磁カロリメータとハドロンカロリメータで構成される。粒子のエネルギーや向きの測定を行う。

- 電磁カロリメータ

バレル部が $|\eta| < 1.475$ 、エンドキャップ部が $1.375 < |\eta| < 3.2$ をカバーする。吸収剤に鉛を用いており、ここで入射した粒子が電磁シャワーと呼ばれる電子や光子の粒子群に変換される [7]。シャワーが液体アルゴンをイオン化した後、これを電極で読み出すことによりエネルギーを測定する。電磁カロリメータは特に光子、電子に適した構造をしている。

- ハドロンカロリメータ

Tile calorimeter、LAr hadronic end-cap calorimeter、LAr forward calorimeter の 3 つから構成される。

Tile calorimeter は電磁カロリメータのバレル部を覆うような位置関係になっており $|\eta| < 1.7$ をカバーする。吸収剤の役割をする鉄とプラスチックシンチレータがタイル状で交互に重なった構造をし、電磁カロリメータで吸収しきれなかったハドロンのエネルギーを測定する。

LAr hadronic end-cap calorimeter は電磁カロリメータのエンドキャップ部の外側に位置しており、 $1.5 < |\eta| < 3.2$ をカバーする。吸収剤の銅板と、電磁カロリメータと同じく液体アルゴンで構成される。

LAr forward calorimeter はエンドキャップの中でもビーム軸の近くに位置しており $3.1 < |\eta| < 4.9$ をカバーする。このカロリメータは 3 層で構成されており最も衝突点に近い 1 層目は電磁カロリメータの役割をしている。この層では電磁カロリメータと同じく銅と液体アルゴンを用いている。2、3 層目はハドロンカロリメータの役割をしており吸収剤はタングステンを使用している。

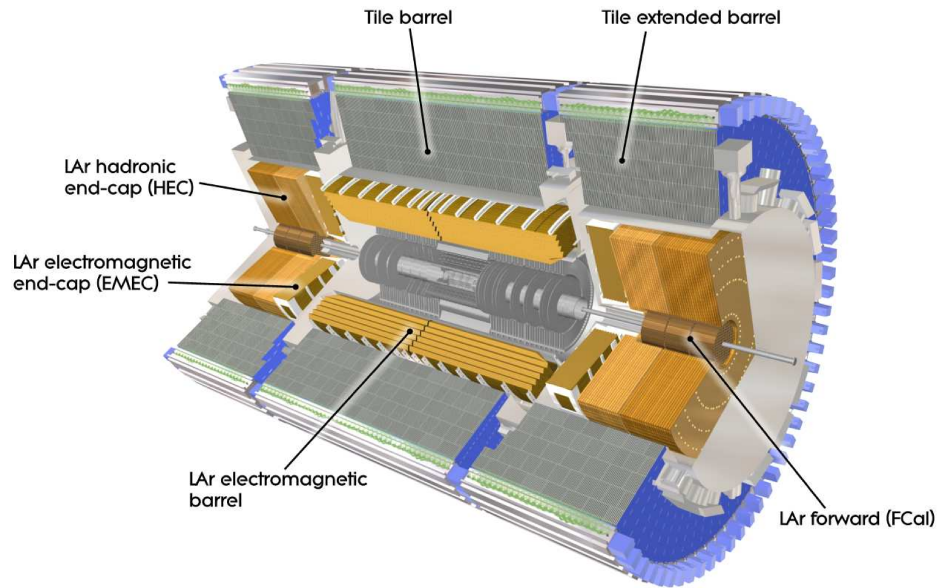


図 2.5: カロリメータ概略図 [5]

2.2.4 ミューオンスペクトロメータ

透過力が高いミュオンの測定を行う。ミュオンの飛跡を曲げるための超電導磁石と飛跡検出器の Monitored Drift Tubes (MDT)、Cathode Strip Chambers (CSC)、トリガーの Resistive Plate Chambers (RPC)、Thin Gap Chambers (TGC) で構成される。それぞれ 2 種類ずつ存在するのはバレル部とエンドキャップ部に分かれているためで、飛跡検出器は $|\eta| < 2.7$ 、トリガーは $|\eta| < 2.0$ をカバーしている。内部飛跡検出器同様に磁場で曲げられたミュオンの飛跡から運動量を測定することが可能となる。そのほか Run3 では TGC の一部が New Small Wheel (NSW) と呼ばれる、より精度の高いものに取り換えられるなど改善がなされた [8]。

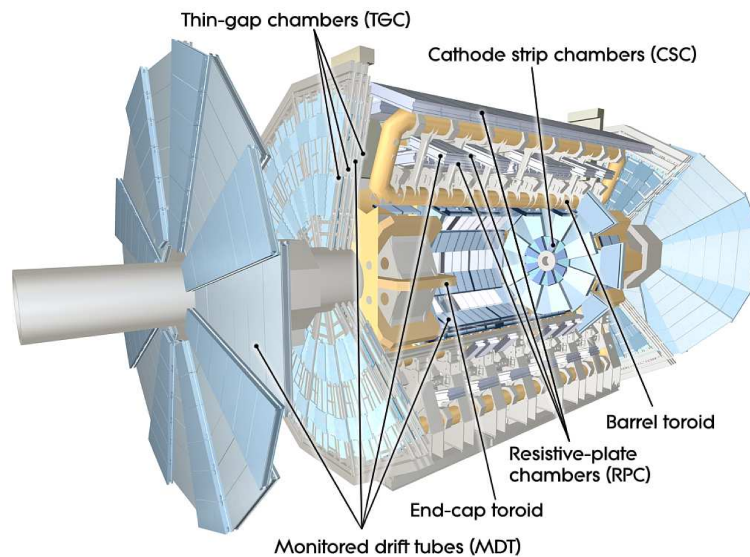


図 2.6: ミューオンスペクトロメータ概略図 [5]

2.3 トリガーシステム

本研究では LHC-ATLAS 実験の Run2 に合わせたシミュレーションデータを使用するため、Run2 におけるトリガーシステムについて記述する。LHC では 40MHz の高頻度で陽子・陽子衝突を起こすが、これを全て記録することは技術的に難しい。そのため重要な衝突事象を選別、記録するトリガーシステムを使用している。トリガーシステムはハードウェアベースの Level-1 Trigger (L1 Trigger) とソフトウェアベースの High Level Trigger (HLT) を用いた 2 段階構成となっている。

- L1 Trigger

L1 Trigger ではカロリメータとミュオントリガーから届く、粗く加工した情報を用いて選別を行う [9]。

カロリメータから送られてくる情報は L1 calorimeter (L1Calo) Trigger で扱う。L1Calo 内にある Preprocessor は送られてきたアナログデータをデジタルに変換し、データに対する校正も行うなど前処理をする。Preprocessor を出たデータは Cluster Processor (CP) と Jet Processor (JEP) に送られる。CP の役割は電子、光子、タウ粒子などの候補を発見することである。JEP の役割は後述するジェットの候補の発見や、ビーム方向に対して横向きエネルギーを足し合わせ、欠損した横方向のエネルギーを計算することである。一方ミュオントリガーである RPC と TGC のヒット情報は L1Muon で扱う。ミュオンの運動量を無限と仮定した場合の通過点と実際の通過点の偏差からミュオンの p_T を推測、6 通りの閾値を設けこれを分類する。L1 Trigger の状態に応じて閾値の設定を使い分ける [10]。

L1 Trigger は $2.5\mu\text{s}$ の間に読み出しを行いデータを記録するか判断することで、40MHz の衝突事象を 100kHz まで削減する。L1 Trigger を通ったイベントは HLT で使用する Regions-of-Interest (RoI, 検出器の η, ϕ の情報) と共に HLT へ送られる。

- HLT Trigger

ソフトウェアによる解析でデータの選別を行う。早さを重視したアルゴリズムによりイベントを削った後、CPU に負荷のかかるアルゴリズムを用いて揺り分ける。検出器に残された情報から、測定した粒子の情報を計算することを再構成と呼ぶが、HLT の段階でも簡易的に再構成を行う。HLT Trigger によってデータは平均 1.2GHz まで削減される。HLT Trigger を通ったデータは永久記憶装置に移され、より時間をかけて再構成が行われる。

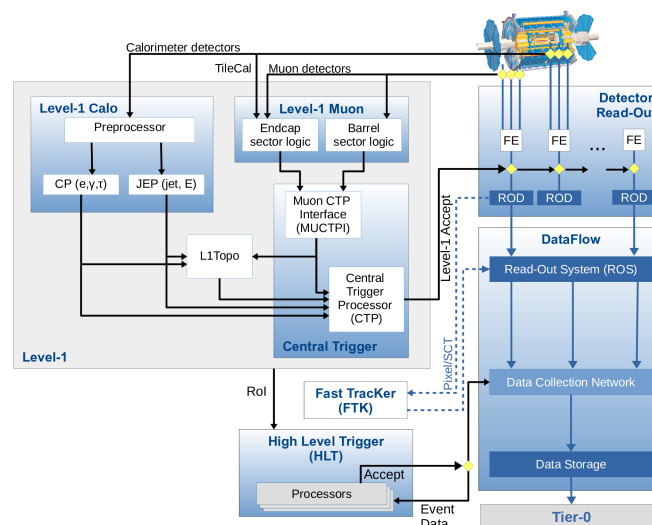


図 2.7: LHC-ATLAS 実験 Run2 におけるトリガーシステム [9]

3 タウ粒子

本研究ではタウ粒子を主に扱う。本章ではタウ粒子の性質や ATLAS 実験における再構成方法、同定手法などを説明する。

3.1 性質

タウ粒子は $-e$ の電荷を持つ第三世代のレプトンであり、質量 $1776.93 \pm 0.93 \text{ MeV}$ とレプトンの中では重い質量を持つ [2]。寿命は $(290.3 \pm 0.5) \times 10^{-15} \text{ s}$ と他のレプトンに比べ短い。崩壊の仕方はレプトン崩壊とハドロン崩壊の大きく 2 つに分けられる。

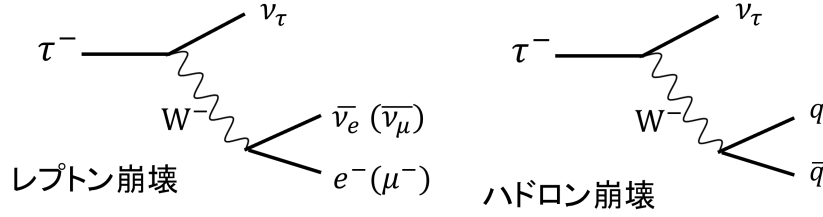


図 3.1: タウ粒子崩壊のファインマンダイアグラム

レプトン崩壊ではタウ粒子よりも軽い荷電レプトンとニュートリノ 2 つを生じる崩壊である。崩壊分岐比は全体の約 35% を占める。残った 65% 程度を占めるのがハドロン崩壊であり、タウ粒子がクォーク対へ崩壊することでハドロンを生じる。電荷の保存則より、終状態に含まれる荷電ハドロンの数も奇数個になることが要求される。この数が 1 つであるものを 1prong、3 つのものを 3prong... と呼ぶ。割合は 1prong が全体の 50%、3prong が 15% 程度を占め、5prong 以降の崩壊が起こる確率は総和が 1% に満たないほど低い。表 3.1 に代表的な崩壊モードを示す。

表 3.1: タウ粒子の代表的な崩壊モード		
崩壊の種類	終状態	崩壊分岐比 [%]
レプトン崩壊	$e^- \bar{\nu}_e \nu_\tau$	17.8
	$\mu^- \bar{\nu}_\mu \nu_\tau$	17.4
ハドロン崩壊		
1prong	$\pi^- \pi^0 \nu_\tau$	25.5
	$\pi^- \nu_\tau$	10.8
	$\pi^- 2\pi^0 \nu_\tau$	9.3
3prong	$\pi^- \pi^+ \pi^- \nu_\tau$	9.3
	$\pi^- \pi^+ \pi^- \pi^0 \nu_\tau$	4.6

3.2 タウ粒子を使用した解析

タウ粒子は精密測定や新物理探索などに広く用いられている。ここではタウ粒子が重要となる解析の例を紹介する。

3.2.1 ヒッグス自己結合定数の測定

標準模型の特徴的な性質として電弱対称性の自発的な破れがある。これはヒッグス場の構造により引き起こされるものであり、ヒッグス自己結合定数がその形の決定に関わる。

ヒッグス自己結合定数の測定はヒッグス三重結合を経由するチャンネルの観測によって行う。これはヒッグス粒子を2つ生じる過程であり、Run2 データを使い $b\bar{b}b\bar{b}$ 、 $b\bar{b}\tau^+\tau^-$ 、 $b\bar{b}\gamma\gamma$ の崩壊チャンネルを利用した測定が LHC-ATLAS 実験で行われている [11]。ヒッグスを2つ生じるファインマンダイアグラムの内、ヒッグス三重結合が含まれるものを図 3.2 に示す。ここで図中の κ_t はトップクォークとヒッグス粒子の結合定数、 κ_λ はヒッグス自己結合定数、 κ_V がヒッグス粒子と、W ボソンまたは Z ボソン (まとめて Vector Boson と呼ばれる) の結合定数である。

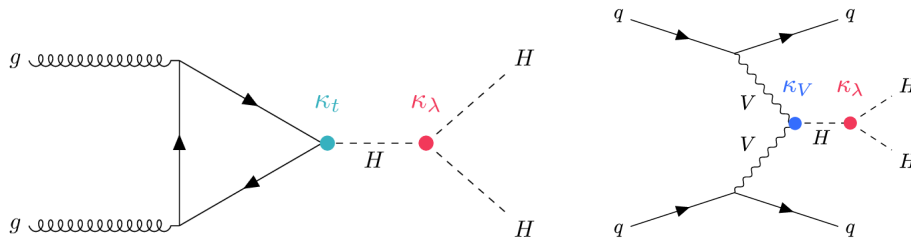


図 3.2: ヒッグス対生成の起こるファインマンダイアグラムの例 [11]

崩壊チャンネルごとに結合の強さ μ_{HH} を求めた結果を図 3.2 に示す。

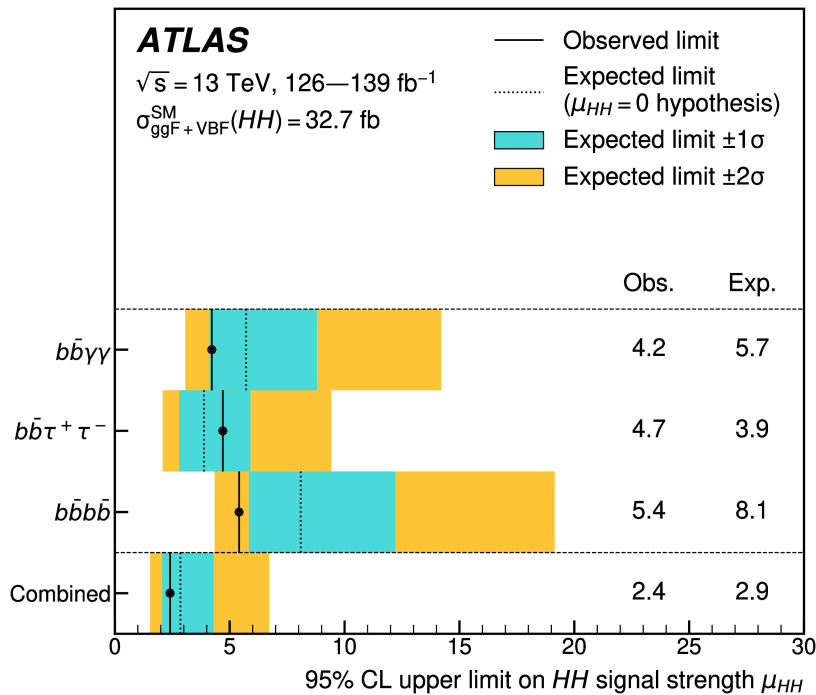


図 3.3: μ_{HH} の予測値と測定値 [11]

複数のチャンネルを組み合わせた μ_{HH} の上限は信頼度 95% で 2.4 をとる。

3.2.2 Heavy Higgs の探索

標準模型をこえる理論の中ではヒッグスセクターを拡張したものがあり、代表例として two-Higgs-doublet models (2HDMs) がある。このモデル群では標準模型で一つだったヒッグス粒子に加え、荷電ヒッグス ($H^\pm$) と重い中性ヒッグス (H/A) が追加される。2HDM の中でも Minimal Supersymmetric SM (MSSM) に代表される typeII と呼ばれるモデルでは、中性ヒッグスからタウ粒子やボトムクォークへの崩壊分岐比が大きくなる。特に MSSM では、標準模型の変数に加え中性ヒッグスの質量 m_A と真空期待値の比 $\tan\beta$ が追加されており、この $\tan\beta$ が大きい場合にタウ粒子対への崩壊分岐比が大きい。LHC-ATLAS 実験では 139fb^{-1} の Run2 データを使用し、 $\tau_{\text{lep}}\tau_{\text{had}}$ 、 $\tau_{\text{had}}\tau_{\text{had}}$ の崩壊モードによる中性ヒッグスの探索が行われた [12]。ここで τ_{lep} 、 τ_{had} はそれぞれタウ粒子のレプトン崩壊、ハドロン崩壊を指す。この解析では $Z \rightarrow \tau\tau$ のような事象と $H \rightarrow \tau\tau$ を区別する必要がある。もしタウ粒子対へ崩壊する前の質量を利用することができればこの分類に有効であるが、タウ粒子は崩壊後にニュートリノを生じるため質量の再構成が難しい。そこで、消失横エネルギー E_T^{miss} を利用した total transverse mass: m_T^{tot} が使用される。

$$m_T^{\text{tot}} = \sqrt{(p_T^{\tau_1} + p_T^{\tau_2} + E_T^{\text{miss}})^2 - (p_T^{\tau_1} + p_T^{\tau_2} + E_T^{\text{miss}})^2} \quad (3.1)$$

ここで τ_1 、 τ_2 は再構成されたタウ粒子 2 つを指す。図 3.4 に $\tau_{\text{lep}}\tau_{\text{had}}$ チャンネルと b-tag を使用した場合の m_T^{tot} 分布を示す。ほかにも $\tau_{\text{had}}\tau_{\text{had}}$ や b-veto を利用したチャンネルによる解析が行われたが、標準模型からの有意な差は認められなかった。

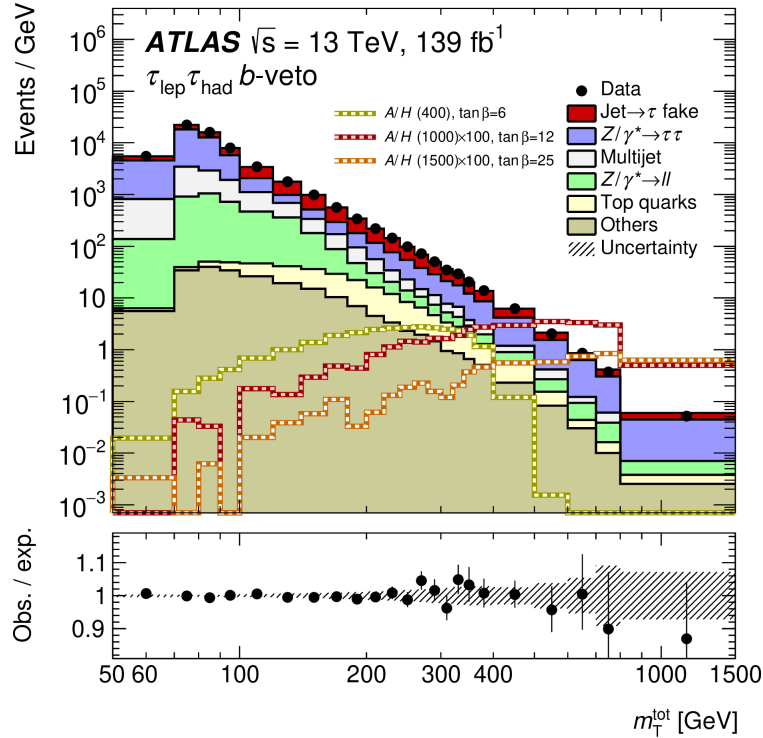
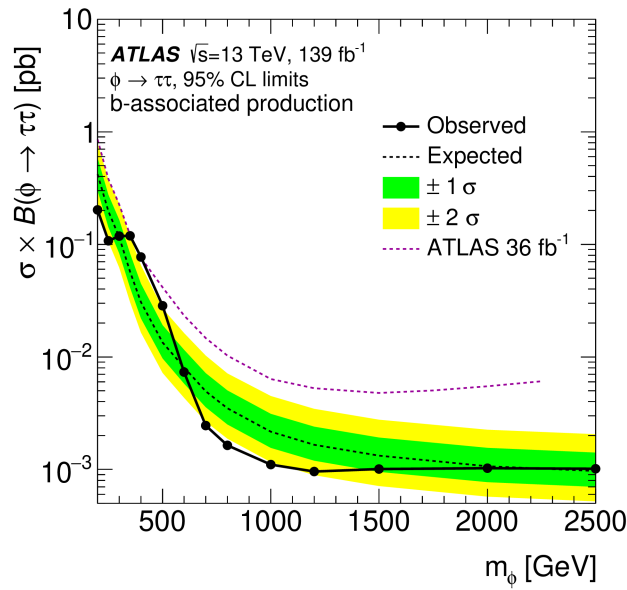


図 3.4: $\tau_{\text{lep}}\tau_{\text{had}}$ と b-tag を使用した場合の m_T^{tot} 分布 [12]

この解析により、重い中性ヒッグスからタウ粒子対へ崩壊する断面積の新たな上限が設けられている。図 3.5 に b-tag を使用した解析から求められた上限を示す。

図 3.5: $\phi(h/H/A) \rightarrow \tau\tau$ の断面積に対する上限 [12]

3.2.3 スタウ探索

標準模型を拡張した理論に超対称性理論がある。これは既存の標準模型に対しヒッグスセクターを拡張し、更に標準模型の粒子と対応する、スピンの 1/2 異なる超対称性粒子を導入したものである。中でも軽い超対称性粒子が暗黒物質の候補であったり、超対称性理論によって力の大統一が可能になるなど期待されている理論である。

タウ粒子に対応した超対称性粒子をスタウと呼び、スタウはタウ粒子に崩壊可能であると予測されている。LHC-ATLAS 実験では Run2 データ 139fb^{-1} を使用し、タウ粒子を利用したスタウ探索が行われている [13]。ターゲットとなった事象を図 3.6 に示す。スタウの崩壊後にタウ粒子とニュートラリーノ $\tilde{\chi}_1^0$ が生じると考えられているが、この内 ATLAS 検出器で測定可能なのはタウ粒子のみである。

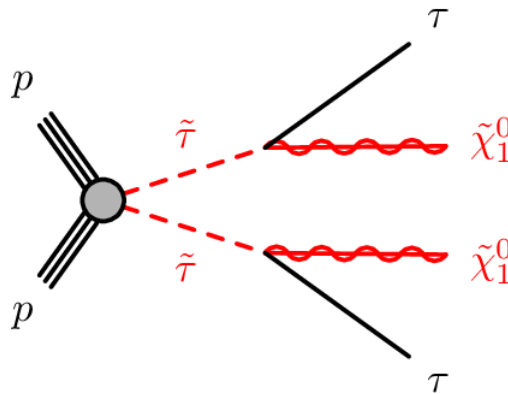


図 3.6: スタウ探索に利用される崩壊事象 [13]

この事象の探索では 2 つのタウ粒子が生じたイベントを使用するが、 $t\bar{t}$ や WW が含まれるイベントではタウ粒子を 2 つ生じる場合があり、それらと区別する必要がある。そのために stransverse

mass: m_{T2} が使用される。

$$m_{T2} = \min_{q_T} \left[\max \left(m_{T,\tau_1}(p_{T,\tau_1}, q_T), m_{T,\tau_2}(p_{T,\tau_2}, p_T^{\text{miss}} - q_T) \right) \right] \quad (3.2)$$

ここで $m_T(p_T, q_T) = \sqrt{2(p_T q_T - p_T \cdot q_T)}$ であり、 q_T は $m_{T,\tau_1}(p_{T,\tau_1}, q_T)$ 、 $m_{T,\tau_2}(p_{T,\tau_2}, q_T)$ が最も小さくなるように選ばれる。 p_T^{miss} は2つのタウ粒子で運動量のベクトル和をとった後、横方向の成分を取り出しマイナスを乗じたものである。 m_{T2} は消失横エネルギーが大きいほど大きくなる。スタウとニュートラリーノの質量差が大きい場合に $t\bar{t}$ や WW よりも大きな消失横エネルギーが生じやすいため、両者の区別に有効である。MC とデータでの m_{T2} 分布を図 3.7 に示す。multi-jet や W +jet のイベントに対しては実データを利用した補正がかけられている。測定結果は標準模型と誤差の範囲内で一致した。

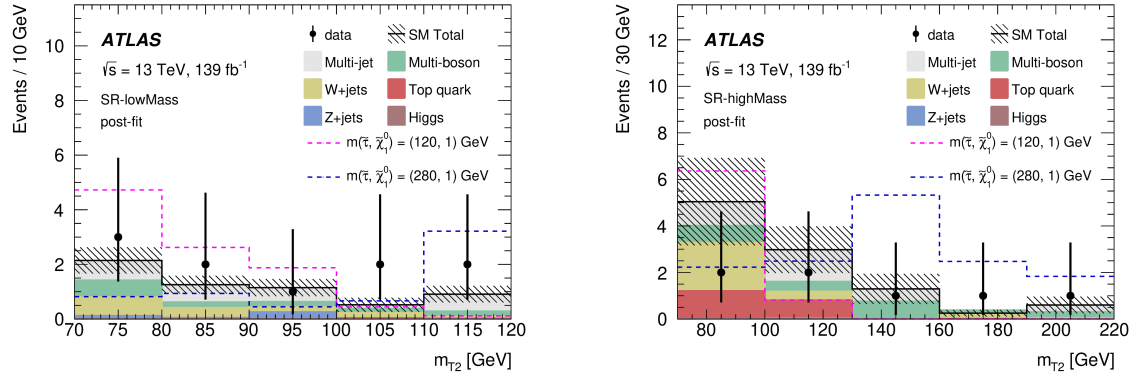


図 3.7: タウ粒子追加の運動量から作成した m_{T2} 分布（左: low mass 領域、右: high mass 領域） [13]

また、本解析によりスタウと最も軽いニュートラリーノの質量に新たな制限が設けられ、スタウの質量が 120GeV から 390GeV の範囲にかけて棄却された。

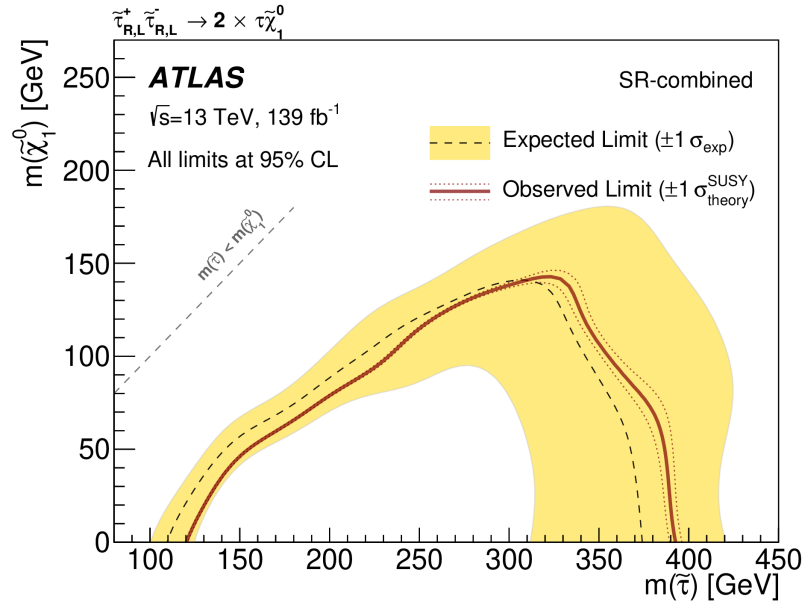


図 3.8: スタウと最も軽いニュートラリーノの質量に付けられた制限 [13]

3.3 LHC-ATLAS 実験におけるタウ粒子の再構成手法

タウ粒子は寿命が短いため、高運動量のものを除き大半が ATLAS 検出器の飛跡検出器最内層へ到達する前に崩壊する。レプトン崩壊の場合はそのまま崩壊後のレプトンとして再構成されるため、タウ粒子として再構成されるのはハドロン崩壊したものである。また、タウ粒子の崩壊ではニュートリノが生じるが ATLAS 検出器でニュートリノをとらえることはできないため、厳密にはタウ粒子崩壊後のジェットが再構成される。LHC-ATLAS 実験ではタウ粒子の再構成後、トラック識別による prong 数の判別、更にタウ粒子同定を行う。本章ではこれらの現行手法について説明する。図 3.9 に全体的な流れをまとめる。

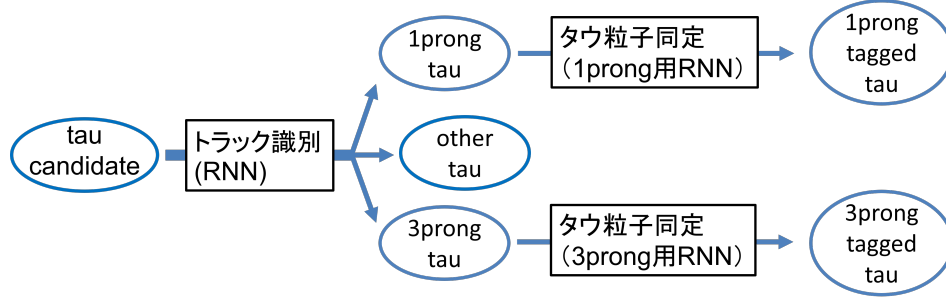


図 3.9: タウ粒子再構成から粒子同定までの手順

最初にタウ粒子の再構成でタウ粒子候補を選び出す。この段階では背景事象の QCD jet が多く含まれている。トラック識別を通して prong 数を判別後、それぞれの prong 数に合わせたタウ粒子同定を行うことで背景事象を取り除く。

3.3.1 タウ粒子再構成

タウ粒子の再構成は、まずジェットを再構成した後に cut base の手法によって行う [14]。ジェットの再構成自体はパラメータ $R=0.4$ とした anti- k_t algorithm を使用する [15]。カロリメータで得たエネルギーの情報はクラスターと呼ばれるオブジェクトとして扱われるが、local hadronic calibration (LC) で補正を行ったクラスターが anti- k_t algorithm に入力される。続いてこのジェットに対し $p_T > 5\text{GeV}$ 、 $|\eta| < 2.5$ のものを選定する。

LHC では複数の生成点が生じるが、ATLAS 実験ではトラックの横運動量の 2 乗和が最大になるようなものを選ぶ。しかしタウ粒子は生成点から少し離れた位置で崩壊する都合上、タウ粒子由来のジェットの生成点は通常の計算方法による生成点から少しずれる。そのため再構成後のタウ粒子に対しては TJVA と呼ばれるアルゴリズムにより生成点を選びなおす [16]。これはジェットの向きに対して $\Delta R < 0.2$ を満たす、ジェットに関連するトラックのみで横運動量和が最大になるよう生成点を取り直すというものである。

3.3.2 トラック識別

トラック識別ではタウ粒子に紐づけられたトラックを 4 種類に分類し、真にタウ粒子から生じたトラックの本数を判定する [14]。

まず再構成されたタウ粒子の運動方向に対し $\Delta R < 0.25$ に収まるトラックのうち以下の条件を満たすものがタウ粒子に紐づけられる（添え字の TJVA がついていない変数は TJVA により生成点を選びなおした後におけるビーム軸から水平方向の距離 d_0 と z 座標 z_0 を指す）。

- タウ粒子の運動方向に対し $\Delta R < 0.25$ に収まる
- $p_T > 1\text{GeV}$
- pixel 検出器に 2 ヒット以上残している
- SCT に 7 ヒット以上残している
- $|d_0^{TJVA}| < 1.0\text{mm}$
- $|z_0^{TJVA} \sin \theta| < 1.5\text{mm}$

$0.25 < \Delta R < 0.4$ のトラックに関しては上記に加え ghost-particle association technique[17] によってタウ粒子候補のジェットとマッチするか、ハドロン崩壊したタウ粒子が最近傍のジェットであった場合に紐づけられる。

次に各トラックに対してトラック識別を行う。トラック識別は深層学習の一種である Recurrent Neural Network (RNN) を利用する。RNN というのは時系列データに対して有効であり、1 つ前のタイムステップの情報を次のタイムステップに使用できる。トラック識別の場合はトラック 1 本 1 本を時系列データとみなしている。例えば 1 番横運動量が高いトラックの情報が 2 番目に横運動量の高いトラックにも使用できる。RNN は様々な派生が作られておりトラック識別に使われているのは Bi-directional Long Short Term Memory (BLSTM) である。

モデルの学習では MC サンプルを使用している。学習に使用するサンプルは信号事象が $\gamma^* \rightarrow \tau\tau$ 、背景事象が dijet であり、それぞれでタウ粒子の横運動量に対し $p_T > 5.5\text{GeV}$ を要求している。モデルへの入力変数を表 3.2 に示す。

分類に使用するモデルはトラックの入力変数が 3 層の BLSTM を通過後、2 度の線形変換を行い最終的に以下の 4 つの track type へ分類する。

- Tau Track (TT)
タウ粒子の崩壊によって生じる $\pi^\pm$ のような 荷電粒子の残すトラック。
- Conversion Track (CT) $\gamma \rightarrow ee$ で生じた電子、陽電子が作るトラック。タウ粒子の崩壊によって生じる π^0 は光子への崩壊分岐比が高いため重要となる。
- Isolation Track (IT) タウ粒子の崩壊とは別の相互作用から生じたジェットのトラック。タウ粒子由来のハドロンとは異なりトラックが重なりやすい、ジェットの中心から外向きへ飛びやすいといった特徴がある。
- Fake Track (FT) 上記に属さないトラック。トラックの再構成に失敗したものやパイルアップによるトラックなどが該当する。

このうち Tau Track の本数がタウ粒子の prong 数として扱われる。タウ粒子同定に進むのは 1prong と 3prong のイベントである。トラック識別の精度を示す混同行列を、truth 1prong tau、truth 3prong tau で分けて図 3.10 に示す。縦軸が正解のクラス、横軸が RNN の判断したクラスを表す。各 track type ごとに規格化されており、prong 数の判別に強く関わる Tau Track の識別精度は 1prong で 95.2%、3prong で 93.1% である。図中右上にある Diagonal efficiency は正解のクラスの数で重みづけした、全トラックに対する正答率を表す。

表 3.2: トラック識別に使用する入力変数 [14]

変数	説明
$\ln(p_T^{track})$	Logarithm of the transverse momentum of the track
$\ln(p_T^{seed-jet})$	Logarithm of the transverse momentum of the seed jet
$\frac{p_T^{track}}{p_T^{seed-jet}}$	Ratio between the transverse momentum of the track and seed jet
η^{track}	Pseudo-rapidity of the track
$z_0^{TJVA} \sin \theta$	Tau vertex longitudinal impact parameter
$\ln(r^{Conv})$	Logarithm of the radius of track conversion
$\tanh(\frac{r^{ConvII}}{500})$	Hyperbolic tangent of the radius of track conversion
$\Delta R(\text{track,seed-jet})$	Radial distance of the track and seed jet axis
$\tanh(\frac{d_0^{TJVA}}{10})$	Hyperbolic tangent of the transverse impact parameter
$\frac{q^{track}}{p_T^{track}}$	Ratio of the transverse momentum and the charge of the track
N_{IBL}	Number of hits in the innermost layer of the pixel detector
$N_{sharedPixel}$	Number of hits in the pixel detector shared by multiple tracks
$N_{sharedSCT}$	Number of hits in the SCT shared by multiple tracks
N_{Pixel}	Number of hits in the pixel detector
N_{TRT}	Number of hits in the TRT
N_{SCT}	Number of hits in the SCT
q^{track}	Charge of the track
P_e^{HT}	Electron probability determined from high-threshold (HT) hits in the TRT, that originate from transition radiation caused by highly relativistic particles traversing the radiator material around each straw

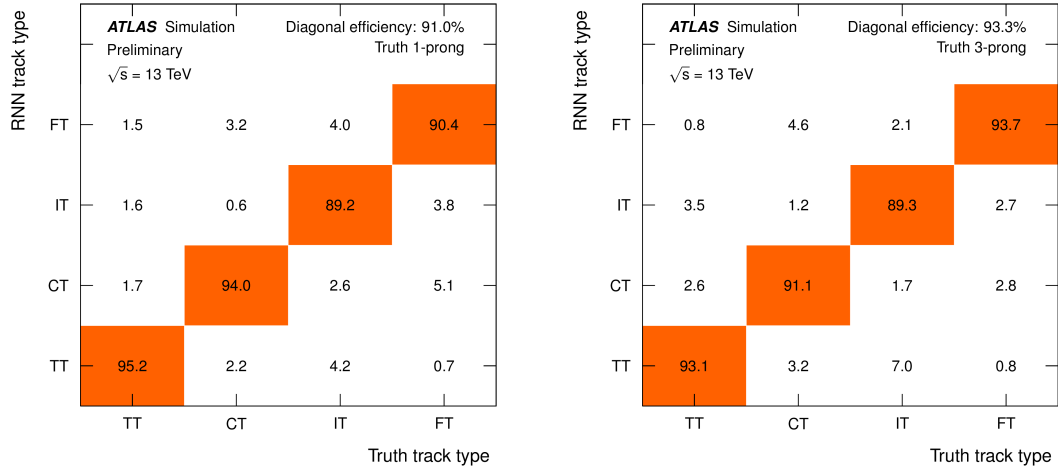


図 3.10: トラック識別の精度 (左: truth 1prong tau、右: truth 3prong tau) [14]

また、truth tau サンプルに対する再構成されたトラックの本数を図 3.11 に示す。truth 1prong tau では 9 割程度の事象に対しタウ粒子由来のトラックを再構成するのに対し、truth 3prong tau では 7 割程度しかトラックを 3 本再構成することができない。

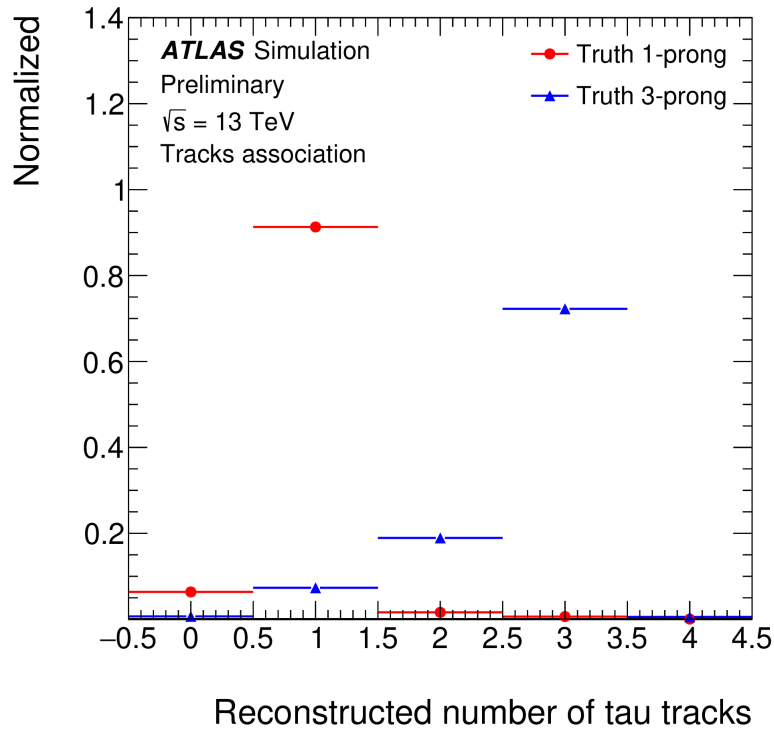


図 3.11: truth tau に対し再構成できたタウ粒子由来のトラックの本数 [14]

3.3.3 タウ粒子同定

タウ粒子同定は、タウ粒子の再構成やトラック識別で取り除くことのできない背景事象を除くために行う。同定手法は度々改善が加えられており、Run2 及び Run3 では RNN の Long Short-Term Memory(LSTM) を取り入れた手法を用いている [14][18]。モデルの学習では信号事象が出力はタウ

粒子 1 つに対して 1 つの score となる。0 から 1 の値をとり、信号事象に対し 1、背景事象に対し 0 に近い値をとるよう学習する。モデルの学習では MC サンプルを使用している。学習に使用するサンプルは信号事象が $\gamma^* \rightarrow \tau\tau$ 、背景事象が dijet である。それぞれでタウ粒子の横運動量に対し $p_T > 20\text{GeV}$ を要求している。図 3.12 に現行手法のタウ粒子同定手法のモデルを示す。

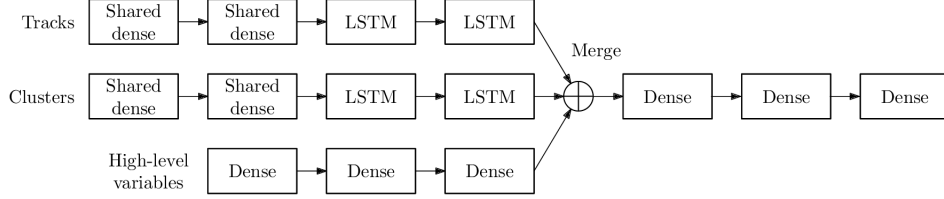


図 3.12: RNN のタウ粒子同定モデル [18]

Dense 層は線形変換を行う層であり、Shared Dense 層では同じく線形変換を行うがパラメータを共有している。また、入力変数は Track、Cluster、そして 2 次情報的な変数が集まった High-level inputs の 3 種類に分けられる。使用されている入力変数を表 3.3 に示す。High-level inputs に関しては 3prong のみで定義される変数があるため 1prong で 8 変数、3prong で 10 変数というように数が異なる。

変数分布の例として High-level inputs の $\Delta R_{max}(track, \tau_{had})$ の分布を図 3.13 に示す。

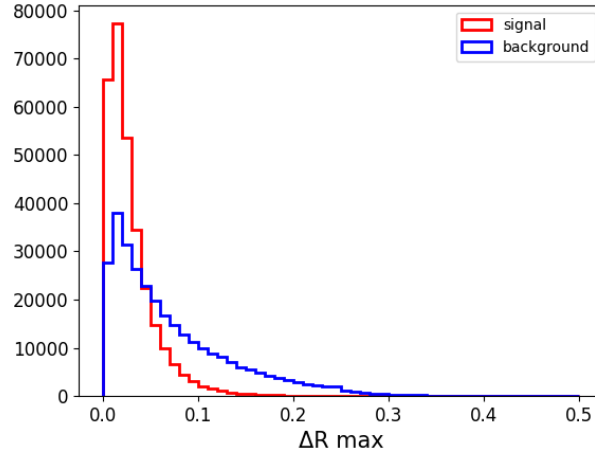


図 3.13: ΔR_{max} の変数分布

粒子が崩壊することによってジェットが生じる場合、崩壊前の粒子の質量が大きくなるほどジェットの広がりには大きくなる傾向にある。これは生成したハドロンが崩壊前の粒子の向きに対して横向きに飛ばされるが、その際に崩壊前の粒子の質量に応じて大きな運動エネルギーを持つためである。タウ粒子の場合は質量が 1.8GeV 程度と、崩壊でハドロンを生じる粒子の中では質量が比較的小さいため、 ΔR_{max} は背景事象より狭い分布になる。このような入力変数の分布の違いを学習することでタウ粒子同定に役立てることができる。

実際にタウ粒子同定をする場合は 1 つ threshold を決め、それより高い score を持つタウ粒子を tag する。一般的に threshold の値を高くすれば、より多くの背景事象を除去できる代わりに tag される信号事象の数は減少する。threshold を推移させた場合の現行手法の性能を図 3.14 に示す。

表 3.3: タウ粒子同定で使用している入力変数 [14]

分類	変数	説明	1prong	3prong
Track inputs	$p_T^{\tau_{had-vis}}$	Transverse momentum of the $\tau_{had-vis}$ candidate	✓	✓
	p_T^{track}	Transverse momentum of each associated track	✓	✓
	$\Delta\eta^{track}$	Angular distance between the track and $\tau_{had-vis}$ axis	✓	✓
	$\Delta\phi^{track}$	Angular distance between the track and $\tau_{had-vis}$ axis	✓	✓
	N_{IBL}	Number of hits in the innermost layer of the pixel detector	✓	✓
	N_{pixel}	Number of hits in the pixel detector	✓	✓
	N_{SCT}	Number of hits in the SCT	✓	✓
	$z_0^{TJV A sin\theta}$	Tau vertex longitudinal impact parameter	✓	✓
	$d_0^{TJV A}$	Tau vertex transverse impact parameter	✓	✓
	$S(z_0^{TJV A sin\theta})$	Tau vertex longitudinal impact parameter divided by its uncertainty	✓	✓
	$S(d_0^{TJV A})$	Tau vertex transverse impact parameter divided by its uncertainty	✓	✓
Cluster inputs	$p_T^{\tau_{had-vis}}$	Transverse momentum of the $\tau_{had-vis}$ candidate	✓	✓
	$E_T^{cluster}$	TopoCluster transverse energy	✓	✓
	$\Delta\eta^{cluster}$	the angular distance between the TopoCluster and the $\tau_{had-vis}$ axis	✓	✓
	$\Delta\phi^{cluster}$	the angular distance between the TopoCluster and the $\tau_{had-vis}$ axis	✓	✓
	$<\lambda_{centre}>$	Mean distance of the cluster shower centre from the calorimeter front face measured along the shower axis	✓	✓
	$<\lambda^2>$	Mean distance of a cell from the shower centre along the shower axis	✓	✓
	$<r^2>$	Second moment in the radial distance of cluster cells from the cluster axis	✓	✓
High-level inputs	$p_T^{\tau_{had-vis}}$	Transverse momentum of the $\tau_{had-vis}$ candidate	✓	✓
	$f_{cent}^{\Delta R < 0.1}$	Fraction of transverse energy at EM scale deposited in the region $\Delta R < 0.1$ with respect to all energy deposited in $\Delta R < 0.2$ around the $\tau_{had-vis}$ axis	✓	✓
	$f_{lead-track}^{-1}$	The transverse energy sum, calibrated at the EM energy scale, deposited in all cells belonging to TopoClusters in the core region, divided by the transverse momentum of the highest- p_T core track of the $\tau_{had-vis}$ candidate	✓	✓
	$\Delta R_{max}(track, \tau_{had})$	The maximum ΔR between a core track associated with the $\tau_{had-vis}$ candidate and the $\tau_{had-vis}$ direction	✓	✓
	S_T^{flight}	The decay length of the secondary vertex in the transverse plane, calculated with respect to the τ vertex, divided by its estimated uncertainty		✓
	f_{track}^{iso}	Scalar sum of the p_T of isolation tracks associated with the $\tau_{had-vis}$ candidate divided by the sum of the p_T of all the associated core and isolation tracks	✓	✓
	f_{track}^{EM}	Ratio of the sum of cluster energy deposited in the electromagnetic part of the TopoClusters (pre-sampler, first and second layers of the LAr calorimeter) to the sum of the momentum of core tracks	✓	✓
	$p_T^{EM+track}/p_T$	Ratio of the $\tau_{had-vis}$ p_T , estimated using the vector sum of core track momenta and up to two most energetic EM clusters in the core region to the calorimeter-only measurement of $\tau_{had-vis}$ p_T	✓	✓
	$m^{EM+track}$	Invariant mass of the system composed of the core tracks and up to two most energetic EM clusters in the core region and the four-momentum of an EM cluster is calculated assuming zero mass and using TopoCluster seed direction	✓	✓
	m^{track}	Invariant mass calculated from the sum of the four-momenta of all core and isolation tracks, assuming a pion mass for each track		✓

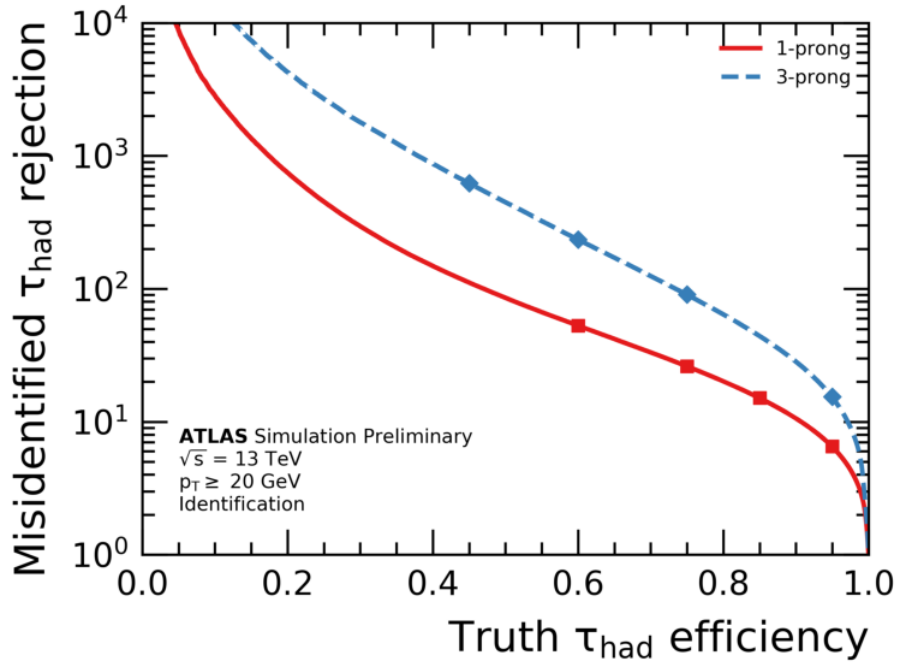


図 3.14: 現行手法でのタウ粒子同定性能 [14]

横軸の Truth τ_{had} efficiency と縦軸の Misidentified τ_{had} rejection は、図 3.9 を使い 1prong を例にして表現すると次のように計算される。

$$\text{Truth } \tau_{had} \text{ efficiency} = \frac{\text{Truth 1prong tau included in "1prong tagged tau"}}{\text{Truth 1prong tau included in "1prong tau"}} \quad (3.3)$$

$$\text{Misidentified } \tau_{had} \text{ rejection} = \frac{\text{Background included in "1prong tagged tau"}}{\text{Background included in "1prong tau"}} \quad (3.4)$$

解析に使用する際は、その目的に合わせて本物のタウ粒子がどの程度残ってほしいか考慮し threshold を選ぶ必要がある。この基準点が 1prong、3prong で 4 つずつ決められており図 3.14 に点が打たれている。各基準点の名前と信号取得効率を表 3.4 に示す。

表 3.4: タウ粒子同定の信号取得効率の基準点

信号事象取得効率		
基準点	1prong	3prong
Tight	60 %	45 %
Medium	75 %	60 %
Loose	85 %	75 %
Very Loose	95 %	95 %

3.4 タウ粒子同定手法の先行研究

タウ粒子同定手法は度々改善がなされており、解析感度向上のためにも同定精度を上げることが重要である。Run1 や Run2 初期のころ、タウ粒子同定には Boosted Decision Tree (BDT) が使用され

ていた [16]、[20]。RNN のようにトラックやクラスターの変数は使わず、ジェットに紐づけられた変数を使用し分類を行う。使用されていた変数は現在の RNN の High-level inputs のものに近く、この時の変数が RNN にも利用されている。2010 年代に深層学習分野で RNN が注目を集めるようになるが、タウ粒子同定手法にも取り入れられた結果、BDT を上回る結果を出したことから RNN に取って代わられた背景がある。このようにアルゴリズムを改善することでタウ粒子同定の精度を上げる試みは今までも行われており、今後も新しい技術の登場に合わせて改善を行うことは解析感度を上げるうえで重要となる。

また、先行研究として昨年の早稲田大学修士卒業生が行った、GNN を用いたタウ粒子同定手法の開発の研究がある [21]。この研究では GraphSAGE という種類の GNN を用いてタウ粒子同定モデルを作成し RNN と性能を比較、prong 数毎に分けてモデルを学習させた場合に RNN の性能を上回ることが示している。

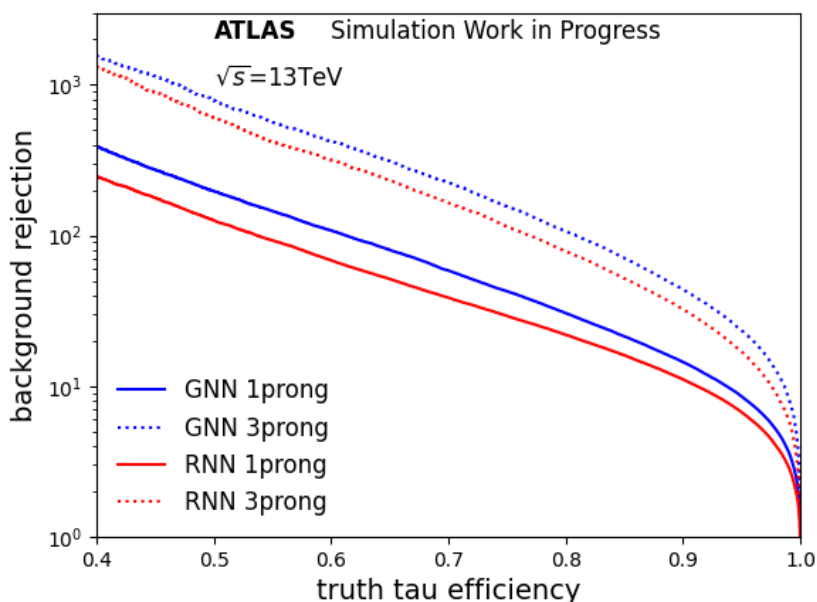


図 3.15: GNN を用いた先行研究での RNN との性能比較 [21]

LHC-ATLAS 実験内では GNN の中でも Graph Attention Network (GAT) を用いたモデルの開発が進められている。b クォークや c クォーク由来の jet を同定する Jet Flavour Tagging において GAT を用いた研究があり [22]、この手法が成果を挙げているためタウ粒子同定手法にも似たモデルを利用しようという研究が現在進められている。

GAT は Attention と呼ばれる機構と GNN を組み合わせたモデルである。Attention によりデータの中から効果的に特徴を抽出することが期待できる。この Attention は本研究で主に扱う Transformer でも取り入れられており近年注目されている技術であるが、重みの計算方法などで GAT とは細部が異なる。

4 本研究で取り入れるタウ粒子同定手法

本研究は新たなタウ粒子同定手法を開発、評価し現行手法からの改善を目標にしている。既存の RNN に代わる深層学習を導入することでこれを旨とする。メインで研究に取り入れるのが Transformer であり、比較対象として昨年度の卒業生が研究していた GNN モデルも使用する。実装の際は PyTorch ライブラリを主に使用している。

4.1 Transformer

Transformer は 2017 年に Google の研究者らから発表された”Attention Is All You Need”[23] で紹介されている深層学習モデルである。当初は自然言語分野で特に目覚ましい結果を出しており、派生されたモデルが画像認識など他の分野でも活用されている。

タウ粒子同定では RNN（中でも LSTM）が使用されていると記述したが、RNN も自然言語や時系列データを扱う上で有効な手法である。RNN の特徴は各系列の特徴を記憶できるという機能にある。これにより時系列データのように順番が意味を持つデータにおいて、ある部分の系列を変換する場合にそれより前の順にあった系列の特徴を有効活用した変換が可能となる。一方で時系列データが長い時、序盤に登場した特徴を終盤まで保持しにくいという問題がある。LSTM は長い時系列データに対しても記憶を保持しやすくするための改良がなされているが、過度に長いと対応できない場合がある。

特徴の抽出を工夫し、より長いデータにも対応できるのが Transformer の強みの 1 つである。一般的な構造を図 4.1 に示す。

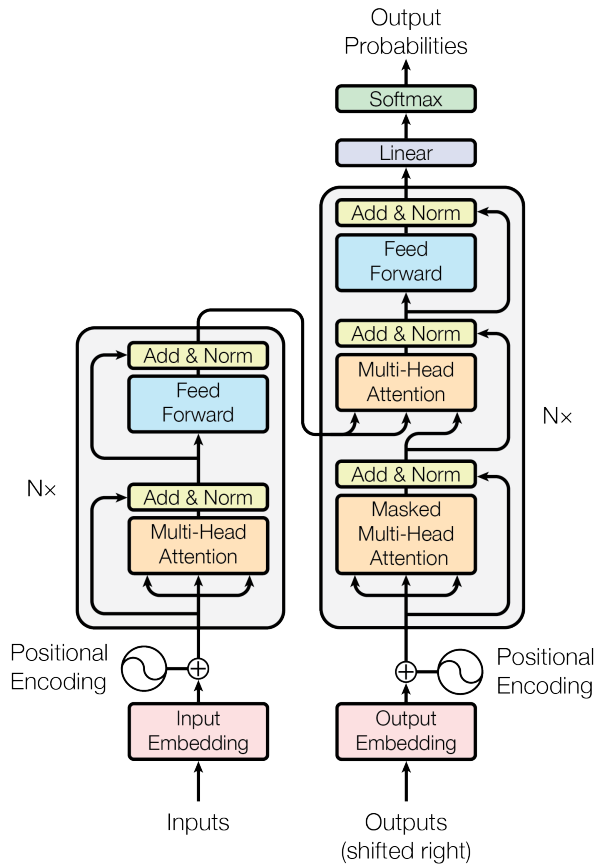


図 4.1: Transformer の構造 (左半分が Encoder、右半分が Decoder を表す) [23]

原著の Transformer は英文をドイツ語へ翻訳するため、英文から特徴を取り出す Encoder と英文に対応するドイツ語の文章を生成する Decoder から構成される。タウ粒子同定で使用するのは Encoder のみである。ここでは Encoder の重要な構成要素である Multi-Head Self-Attention と Feed Forward について触れる。

Multi-Head Self-Attention

Multi-Head Self-Attention は入力の中から強い特徴を探し出す役割を持つ。具体的には関連度が高い特徴量に対して強く重みづけして和をとる。この部分が RNN からの改善点にあたる。Transformer ではこの Multi-Head Self-Attention で最初に各系列の情報を相互に伝搬し合う。リレー形式で伝搬する RNN と異なり、間に別の系列を挟むようなことが無いため離れた系列の情報も利用しやすい。図 4.2 にその構造を示す。

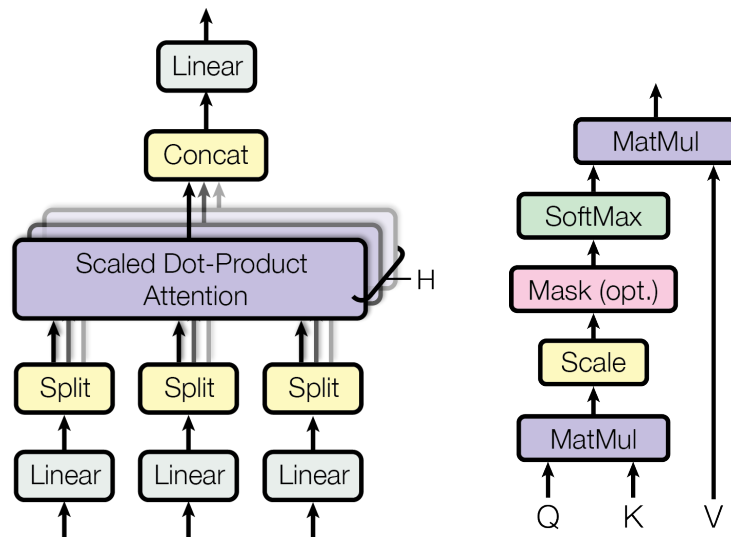


図 4.2: Multi-Head Self-Attention の構造 (右は Scaled Dot-Product Attention の中身)[23]

入力されたデータは独立した線形変換で 3 通りに変換された後、各々 H 枚ずつに分割される。分割することにより複数の特徴に合わせた重みづけができる。この分割する数はヘッドの数などと呼ばれることが多い。

分割された行列は 3 枚組でそれぞれ Scaled Dot-Product Attention に入り、内 2 つで行列積をとった後スケーリングされ、softmax 関数で変換される。この行列が入力されたデータ間の関連度合いを示す。これと 3 つ目の行列で積をとることにより、特徴の関連度合いを加味した重みづけを行う。最後に Scaled Dot-Product Attention を出た行列 H 枚をつなぎ合わせて 1 つの行列に戻した後、線形変換を行う。

Feed Forward

Feed Forward は 2 度の線形変換の間に活性化関数である relu 関数を挟んだ構造になっている。relu 関数は入力した値が 0 以下の時に 0 を返し、0 以上の場合はその値をそのまま返す関数である [24]。線形関数の間に活性化関数を挟む構造自体は Transformer 以前から深層学習の分野で広く用いられおり、脳にあるニューロンの働きを模している。

実装の際は PyTorch ライブラリにある TransformerEncoder 等の関数を使用している [25]。

4.2 GNN

Graph Neural Network (GNN) は頂点（ノード）とそれらをつなぐ辺（エッジ）で構成されるグラフデータに対して有効な深層学習である。各ノードが入力変数を持ち、関係のあるノード間でエッジをつなぐ。

GNN の活用例としてリンク予測やグラフ分類がある。リンク予測というのはノード間のつながりを予測するもので、SNS における個人のつながりなどを予測するタスクが該当する。グラフ分類はノードやエッジの情報からグラフの属性进行分类するというもので、化学物質の性質予測などに利用される。タウ粒子同定もこのグラフ分類である。

本研究では GNN の中でも GraphSAGE と呼ばれる手法を使用する [26]。

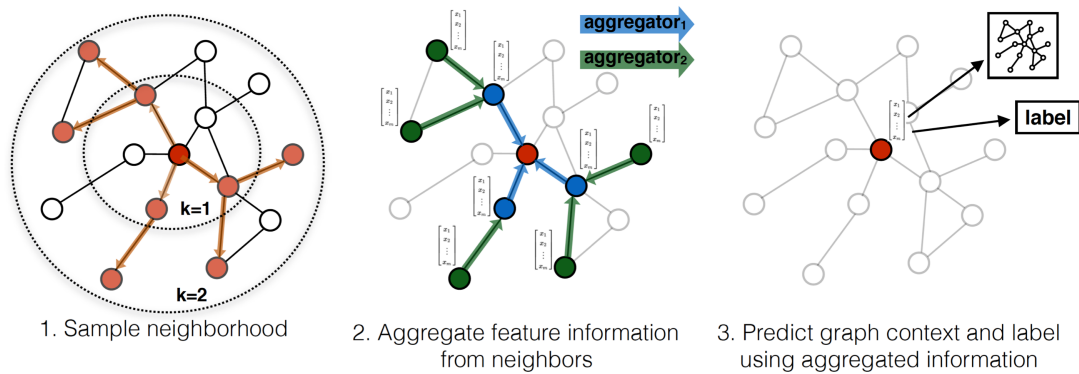


図 4.3: GraphSAGE の計算手法 [26]

GraphSAGE というのはあるノードを変換する際、近傍ノードの情報も集約関数で伝搬させ利用するというものである。本研究では torch_geometric ライブラリの SAGEConv を実装に使用しており、集約関数は周辺ノードの平均をとっている。タウ粒子同定ではノードにトラックやクラスターの特徴量を割り当て、エッジは全てのノードをつなぎ全結合にしている。

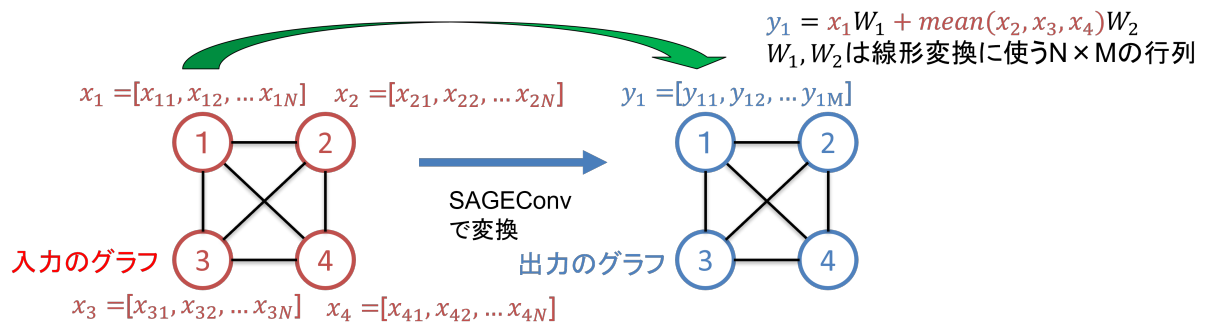


図 4.4: SAGEConv のイメージ図

5 Transformer と現行手法の比較

本章では Transformer 及び GNN のモデルで現行手法と同じく 1prong、3prong それぞれに特化して学習を行い性能を比較する。

5.1 学習で使したサンプルと条件

RNN と同様に、モデルの学習は信号事象が $\gamma^* \rightarrow \tau\tau$ 、背景事象が dijet の MC サンプルを使用する。それぞれでタウ粒子の横運動量に対し $p_T > 20\text{GeV}$ を要求している。モデルを学習する際は 1prong、3prong でそれぞれ分けて行う。1prong 用のモデルを学習する際はトラック識別で 1prong と判断されたサンプルを使用する。更に信号事象に関しては truth の prong 数も 1prong であることを条件に課す。3prong の場合でも同様に prong 数で条件を課す。入力変数は現行手法と同じく表 3.3 で示したものを使用している。

5.2 モデルの構造

学習に使用した Transformer と GNN のモデルをそれぞれ図 5.1、5.2 に示す。

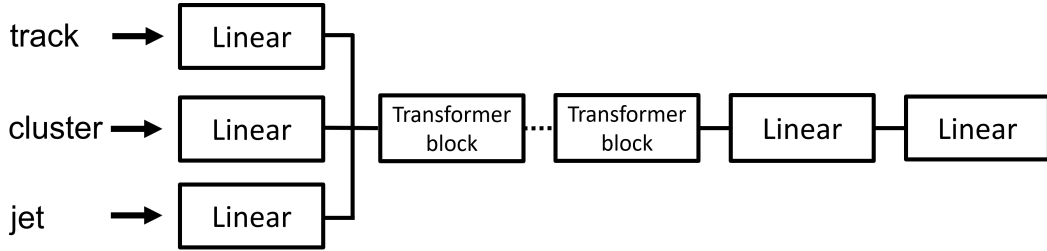


図 5.1: Transformer を使用したタウ粒子同定モデル

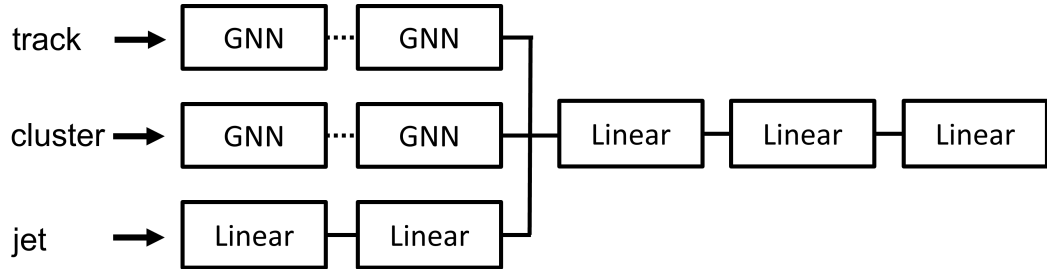


図 5.2: GNN を使用したタウ粒子同定モデル

ここで図中の Linear では線形変換を行っている。Transformer Block は Transformer Encoder を指し、GNN は SAGEConv を指す。また、High-level inputs はハドロン崩壊したタウ粒子ごとに用意される変数であるため図中では jet と表記している。

タウ粒子に紐づけられるトラックやクラスターは複数存在するため、ここではトラックを横運動量が大きい順に最大で 10 本、クラスターを横エネルギーが大きい順に最大で 6 本使用する。この数は現行手法の RNN モデルと同じである。

Transformer モデルでは線形変換により入力変数の次元数をトラック、クラスター、ジェットでそろえた後、行列を連結し 1 つにしている。その後 Transformer Block4 層または 8 層で演算を行い、最

後に線形変換を2度行うことで score の算出を行う。Transformer Block の次元数は 64 であり、ヘッドの数は 4 にしている。この次元数と層数は 1prong 用のモデルで検証した結果選んだ値であるが、この内容は Appendix に回す。

GNN モデルではトラック、クラスターの入力変数は 4 層の GNN Layer に入り、ジェットの入力変数は 2 層の線形層で変換を行う。GNN Layer の次元数は 512 である。その後、行列を連結し線形変換することで score を求める。

5.3 モデルの学習方法

深層学習では一般的に train、validation、test の 3 種類にデータサンプルを分けモデルの学習や評価を行う。本研究でもこの形をとる。

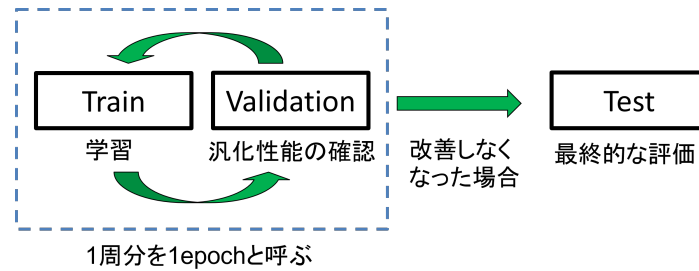


図 5.3: 学習から評価までの流れ

モデルは train data に合わせてパラメータを調整する。しかし、train data だけでは調整の仕方が偏ってしまい他のデータセットに対する汎用性が損なわれてしまうことがある。これを避けるため、validation data に対しても改善が見られるか検証することで汎化性能を確認する。このサイクル 1 周分を 1epoch と呼ぶ。何 epoch か進んだ後、validation data で改善が見られなくなった段階で学習を打ち切り test data で最終的なモデルの性能を評価する。このようにデータを 3 つに分け独立させることでバイアスの少ない学習、評価を行う。本章で使用したサンプルの大きさを表 5.1 に示す。

表 5.1: タウ粒子同定モデルの学習に使用した事象数

種類	Train	Validation	Test
1prong signal	1.6M	0.4M	0.3M
1prong background	1.6M	0.4M	0.3M
3prong signal	0.8M	0.2M	0.3M
3prong background	0.8M	0.2M	0.3M

モデルの学習が進んでいるかどうかは、あらかじめ用意した損失関数 (loss) を用いて判断する。損失関数というのはモデルが出した答えの不正解の度合いを算出する関数であり、学習はこの計算結果が小さくなるように進められる。本研究では損失関数が 15 回連続で最小値を更新しなかった場合に学習を打ち切り、損失関数が最も低かった epoch のパラメータを採用する。タウ粒子同定の場合は Binary Cross Entropy を用いる。

$$loss = -\frac{1}{|N|} \sum_{n \in N} w_n (y_n \log x_n + (1 - y_n) \log(1 - x_n)) \quad (5.1)$$

ここで N はサンプルの集合を示し、 x_n が n 番目の事象に対しモデルが算出した score、 y_n がラベルを指す。ラベルはその事象がタウ粒子であった場合に 1、タウ粒子でない場合に 0 をとる。 w_n は各事象での重みづけを表す。

重みづけはタウ粒子の横運動量で決める。図 5.4 に信号事象と背景事象での横運動量分布を示す。

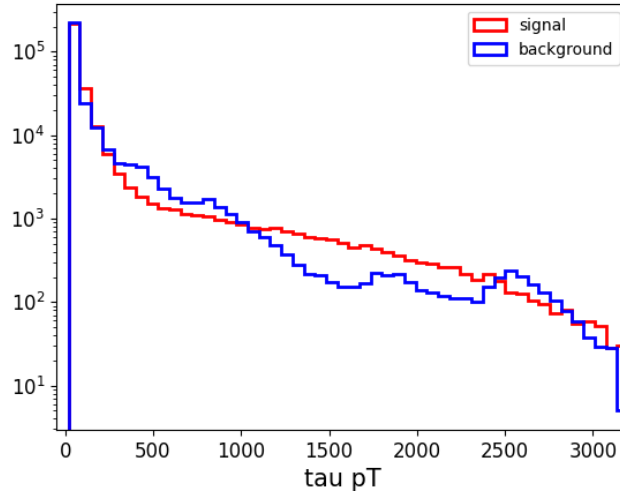


図 5.4: 信号事象と背景事象での横運動量分布

横運動量の大きさによって信号事象と背景事象の分布の仕方が異なる部分がある。このまま学習を行った場合、単に信号事象の多い領域で score を高くし、信号事象の低い領域で score を低くするように学習をしてしまう可能性がある。このような意図しないバイアスを避けるため重みづけを行い学習する。具体的には背景事象に対してのみ $w_n = \frac{\text{信号事象の数}}{\text{背景事象の数}}$ として、信号事象は $w_n = 1$ で重みづけを行う。

5.4 性能比較

各モデルの背景事象除去能力を図 5.5 に示す。図中の縦軸、横軸は 3.14 の Misidentified τ_{had} rejection、Truth τ_{had} efficiency と名前は異なっているが同じ方法で計算している。また、RNN による曲線に関しては LHC-ATLAS 実験グループで使われている RNN モデルから算出された score をもとに作成したものであり、自分で学習させたものではない。

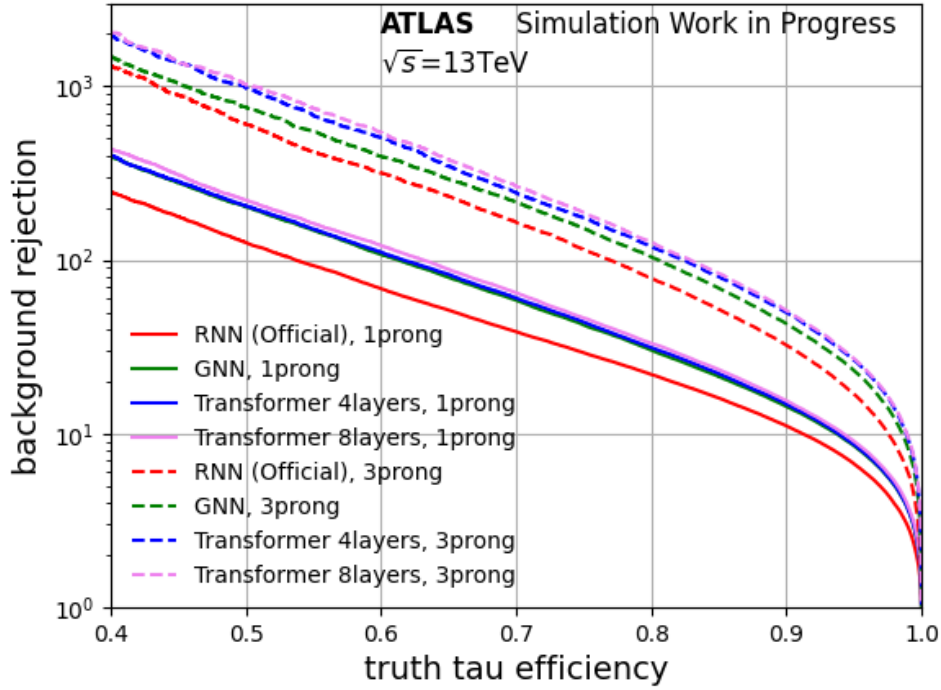


図 5.5: 各モデルの背景事象除去能力

また、各基準点での背景事象除去能力を表 5.2 に示す。

表 5.2: 各基準点での背景事象除去能力

モデルの種類	背景事象除去能力		
	Loose	Medium	Tight
RNN, 1prong	16.1	29.3	68.5
GNN, 1prong	21.3	42.3	107.5
Transformer 4 層, 1prong	22.2	44.0	111.0
Transformer 8 層, 1prong	23.4	46.4	121.1
RNN, 3prong	114.8	316.5	890.2
GNN, 3prong	152.5	393.7	1038.1
Transformer 4 層, 3prong	174.9	506.8	1369.9
Transformer 8 層, 3prong	187.0	540.5	1376.1

図 5.5 で背景事象除去能力を比較すると、信号事象取得効率によらず Transformer が RNN や GNN よりも優れていることがわかる。表 5.2 で信号取得効率が 60% である 1prong の Tight、3prong の Medium を確認すると、RNN から Transformer 8 層で背景事象除去能力が 1.7 倍以上大きくなることがわかる。また、Transformer を 4 層にした場合よりも 8 層にしたほうが除去能力は高く、層数を増やすことでの性能向上が見られる。

6 タウ粒子を崩壊後の prong 数で分けずに学習を行った Transformer モデル

今まで扱っていた Transformer ではトラック識別で 1prong と判別されたサンプルと 3prong と判別されたサンプルで分け、それぞれに特化したモデルを作成していた。しかし、一般的に深層学習を行う上ではデータをまとめて使い大規模な学習を行った方がよい。そのため本章ではタウ粒子崩壊後の prong 数でサンプルを分けず、ひとまとめにしたサンプルでモデルの学習を行った。この学習用のサンプルにはトラック識別で 0prong や 2prong と判別され、今までタウ粒子同定モデルの学習に使用していなかった事象も含まれている。これにより学習用データのサイズを増やすことができる。また、prong 数に関係なく 1 つのモデルでタウ粒子同定が可能となれば 1prong や 3prong 以外に分類された事象への利用も今後期待できる。

6.1 学習と評価方法

この Transformer での学習と評価手順を図 6.1 に示す。

学習時



評価時

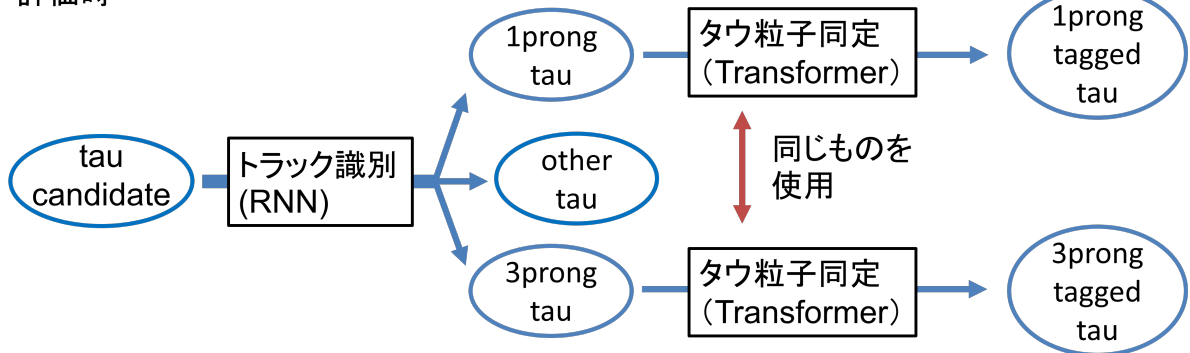


図 6.1: サンプルを prong 数で分けない場合の Transformer の学習と評価手順

学習時は再構成後のタウ粒子に対しタウ粒子同定のみを行いトラック識別は関与しない。そのため、もしトラック識別を行えば 0prong や 2prong と判別されていたであろう事象も含まれている。評価時は RNN との公平な比較のため、トラック識別で 1prong または 3prong と判別された test data を使用することで条件を揃える。使用したサンプルのサイズは、Train が signal、background で 6.4M ずつ、Validation が 1.6M ずつである。signal のサンプルに占める truth 1prong tau と truth 3prong tau の割合は、約 10:3 で 1prong のほうが多くなっている。これはシミュレーションでサンプルの作成を行った際にタウ粒子の崩壊分岐比が反映されているためである。また、test data に関しては前の章同様に prong 数判別後の事象を使用している。そのため 1prong で評価を行う場合は 1prong と判別された truth 1prong tau を signal とし、これと 1prong と判別された background を使用した。また、学習に使用するサンプルの数が増えたことで学習に必要な epoch 数が減少したため、loss が 5 回連続で更新されなかった場合に学習を打ち切るように変更している。

Transformer のモデルは 8 層、次元数を 256 に変更し、ヘッドの数を 4 つにして学習を行った。使用する入力変数は 3prong のものに加え、RNN のトラック識別に使う変数も利用している。3prong の入力変数の中には、ジェットの中心付近のトラックの本数が 3 本存在しなければ計算できないような定義のものがある。一部の事象ではそういった変数を計算できず、初期値など意味のない値が入っている場合もあるがそのまま学習に使用した。また、使用するトラックの本数の上限を 30 本に引き上げている。

6.2 性能比較

RNN と新たに学習させた Transformer での背景事象除去能力を図 6.2 に、各基準点での数値を表 6.1 に示す。凡例にある 1prong、3prong というのは使用している test data の signal における真の prong 数を指している。参考として載せている Transformer, 2prong は、truth 3prong tau かつトラック識別で 2prong と判別された事象に対する Transformer の背景事象除去能力を示している。2prong での Loose や Medium などの基準点は定義されていないため、表 6.1 に載せている値は 3prong と同じ定義を利用している (1prong、3prong での定義は表 3.4 に掲載)。また、前述したように RNN の 1prong、3prong ではそれぞれの prong 数に特化したモデルを使用しているが、今回の Transformer モデルは prong 数に関わらず同じモデルを使用している。

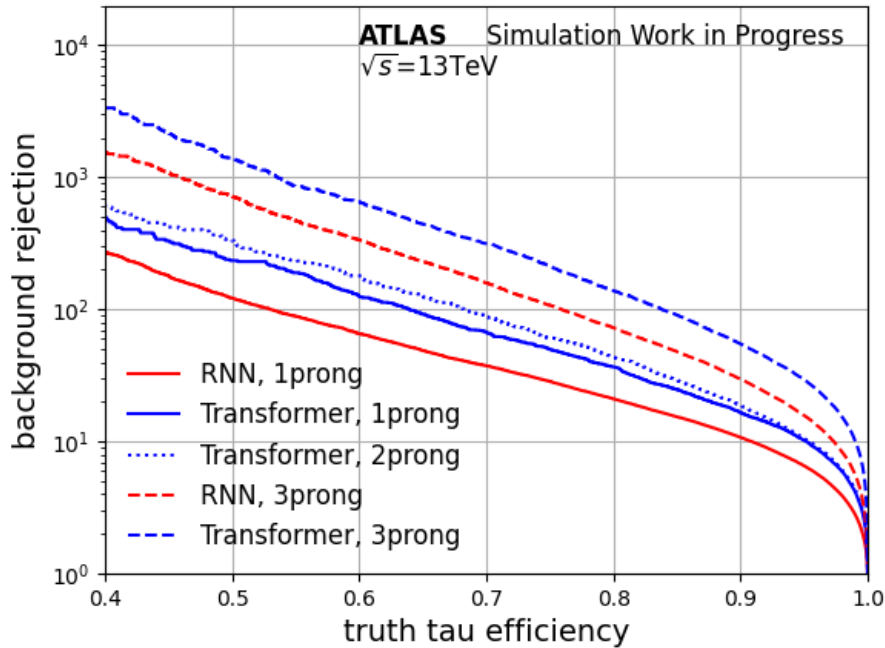


図 6.2: prong 数で分けずに学習させた Transformer モデルと RNN の背景事象除去能力

表 6.1: prong 数で分けずに学習させた Transformer における各基準点での背景事象除去能力

モデルの種類	背景事象除去能力		
	Loose	Medium	Tight
RNN, 1prong	15.4	28.6	65.6
Transformer, 1prong	24.7	49.9	125.5
Transformer, 2prong	61.1	180.0	421.5
RNN, 3prong	107.9	337.5	1086.9
Transformer, 3prong	213.0	654.1	2206.2

前の章と RNN の背景事象除去能力の値が異なるのは使用するサンプルが異なっているためである。基準点での値を見ると Transformer では 1prong、3prong 共にいずれの値も前の章からの向上している。3prong で特に増え幅が大きく、Tight に至っては 1.6 倍程度まで数値が増加している。変更点が多いためどれが根本的な影響を与えているか判断しづらいが、タウ粒子由来のトラックの本数が多い 3prong で改善の度合いが強いことを考えると、トラックの変数を追加したことが性能の向上に繋がった可能性がある。

また参考として載せた 2prong の結果では、truth 3prong tau の事象を使用しているにもかかわらず背景事象除去能力が 3prong から大きく下がっていることがわかる。後で詳しく扱うが、トラック識別で 2prong と判別される信号事象のほとんどが truth 3prong tau かつタウ粒子の飛跡が 2 本しか再構成できなかったものである。このことを鑑みると背景事象除去能力は再構成できたタウ粒子の飛跡の本数に影響を受けると考えられる。しかし 2prong の結果が 3prong のものに劣るとはいえ、1prong の事象と同程度には背景事象を取り除くことができるため、解析への使用が期待できる。そのためには現行手法から prong 数の判別方法を変え、現状 2prong と判別されるような信号事象も正しく 3prong と判別できるようにするなどの改善が必要である。

7 prong 数の判別とタウ粒子同定を行う Transformer モデル

前の章ではサンプルを prong 数で分けないことで 1 つのモデルの学習に使用するデータサイズを増やし、変数やトラックの数も追加することでタウ粒子同定性能を上げた。現状使用できるサンプルを考えると、タウ粒子同定のみでここからさらに性能を劇的に向上させるのは難しい。

そこで本章ではタウ粒子同定の前行われる prong 数の判別方法にも改善を加える。現行手法では RNN によるトラック識別で 1prong または 3prong と判別された事象のみがタウ粒子同定に進み、0prong や 2prong と判別されたタウ粒子はこの時点で除外されてしまう。これを改善するためトラック識別に代わる prong 数の判別方法を Transformer で考案した。prong 数の判別精度を高め truth tau を多く残すことで最終的な信号事象取得効率の向上を目指す。

7.1 使用するモデルと分類方法

本章で新しく扱う Transformer モデルでの同定までの手順を図 7.1 に示す。ATLAS 標準のトラック識別やタウ粒子同定の RNN を使わず、1 つの Transformer モデルで prong 数の判別とタウ粒子同定を両方行う。

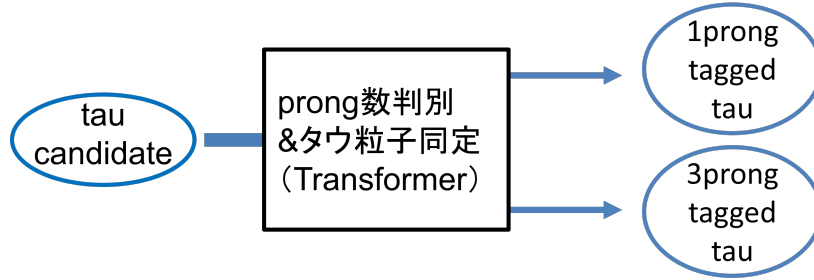


図 7.1: prong 数の判別とタウ粒子同定を行う Transformer モデルの同定手順

使用する Transformer のモデルは前の章と基本的に同じだが、最後の Linear 層で出力の数を 3 つに変える。1 つ目を 1prong tau の score、2 つ目を 3prong tau の score、3 つ目を background の score とした 3 値分類を行うことで prong 数と共にタウ粒子同定の score も判別できるモデルを作成する。分類の様子を図 7.2 に示す。

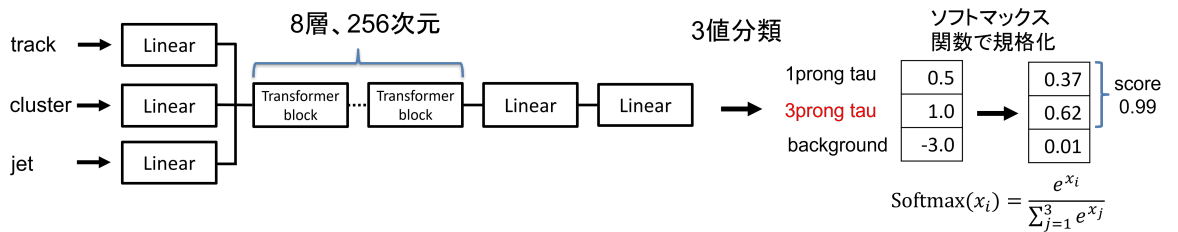


図 7.2: prong 数の判別も行う Transformer

分類の手順を以下に示す。

1. モデルで 3 つの score を算出
2. 1prong tau と 3prong tau の score で比較を行い、大きい方の prong 数をその事象の prong 数として決定

3. softmax 関数で score を規格化

4. 1prong tau score と 3prong tau score を足しタウ粒子同定で使用する score を算出

score の大小を利用して prong 数の判別を行う。現行手法のトラック識別ではタウ粒子に紐づけられたトラックの中から真にタウ粒子由来であると思われるものをラベリングし、その本数を prong 数として扱っていた。そのためラベリングの誤答などから 0prong や 2prong のような事象が生じる場合があり、このような事象は現行手法では解析から除外していた。しかし今回 Transformer で試す手法では 1prong と 3prong のみしか分類先が存在しないため、prong 数の判別で本物のタウ粒子が除外されることはなくなる。なお本章では 5prong 以上の事象も 3prong として扱い学習を行っている。一方、現行手法では 1prong や 3prong と判別されなかった background も除外されるためトラック識別が background の除去に役立つ側面があった。今回の Transformer モデルの場合は background も必ず 1prong か 3prong と分類してしまうため多く残り、background の除去能力はタウ粒子同定頼みとなる。

ここで重要となるのが score の判別方法である。当初は規格化後の score をそのまま利用していたが、この方法だと図 7.2 にあるような 1prong と 3prong で score が割れるケースに対応できない。background score で 0.01 のように低い値をとる事象はタウ粒子として同定しやすい方が自然であるため、1prong と 3prong の score を足して最終的な score を判別するという方法をとっている。各 score 分布は Appendix で掲載する。

またトラック識別の部分を改善する試みは昨年の GNN を使ったタウ粒子同定の研究でも行われており、本研究でも比較対象としてこのモデルを使用する。モデルの構造を図 7.3 に示す。

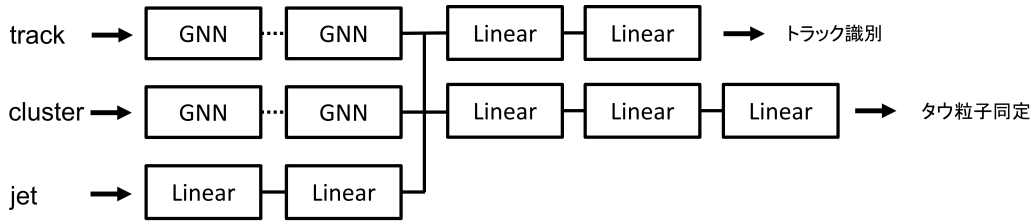


図 7.3: トラック識別とタウ粒子同定を並行して行う GNN モデル

出力が2つに分かれ、1つは今まで同様タウ粒子同定の score、もう1つはトラック識別に対応した score である。トラック識別の方は RNN 同様に、トラック 1 本 1 本に対して 4 値分類を行う。この分類のうち Tau Track (タウ粒子由来の荷電ハドロンが作った飛跡と思しきもの) に分類されたトラックの本数が prong 数となる。そのためこちらでも RNN 同様に 0prong や 2prong と分類される事象が生じる。

Transformer、GNN の新しいモデルではいずれも前の章同様、1prong と 3prong のサンプルを分けてまとめて学習する。GNN ではトラック識別、Transformer では score を 3 つ算出することから多値分類での学習になり、損失関数は以下の Cross Entropy を利用する。

$$loss = -\frac{1}{|N|} \sum_{n \in N} \sum_{c \in C} w_{n,c} y_{n,c} \log x_{n,c} \quad (7.1)$$

ここで N はサンプルの集合である。 C はモデルの分類先の集合を表し、今回の場合では 1prong tau, 3prong tau, background が要素となる。 $x_{n,c}$ は n 番目の事象、 c の分類に対しモデルが算出した score、 $y_{n,c}$ がラベルを指す。 $w_{n,c}$ は各事象での重みづけを表す。今回の場合は学習に使う truth 1prong tau, truth 3prong tau, background の事象数でそれぞれ逆数を取り、これらを規格化したものである。例えば truth 1prong tau が 100 事象、truth 3prong tau が 100 事象、background が 200 事象の場合、weight を 0.4、0.4、0.2 として loss を計算する際に使用する。

7.2 背景事象除去能力の新たな定義

タウ粒子識別だけでなく prong 数の判別方法も変えるため、背景事象除去能力の計算方法を今までのものから変更し評価を行う。図 3.9 を使い 1prong を例にして表現すると次のように計算される。

$$\text{truth tau efficiency} = \frac{\text{Truth 1prong tau included in "1prong tagged tau"}}{\text{Truth 1prong tau included in "tau candidate"}} \quad (7.2)$$

$$\text{Misidentified } \tau_{\text{had}} \text{ rejection} = \frac{\text{Background included in "1prong tagged tau"}}{\text{Background included in "tau candidate"}} \quad (7.3)$$

truth tau efficiency を計算する際に分母の事象数を「1prong と識別されたタウ粒子の数」でなく単に「再構成後のタウ粒子の数」に変えている。同様に background rejection を計算する際も 1prong と認められたタウ粒子の数でなく再構成後のタウ粒子の数を利用している。3prong でも同様である。これにより prong 数の判別能力まで含めたモデルの性能を比較することができる。

7.3 背景事象除去能力の比較

新しい定義の背景事象除去能力を各モデルで比較したものが図 7.4 である。また、各基準点での背景事象除去能力を表 7.1 に示す。ここで新しく学習を行った GNN と Transformer のモデルはそれぞれ名前に inclusive と付けている。また、前章においてタウ粒子の崩壊後の prong 数で分けずに学習を行った Transformer と RNN のトラック識別を使用したものを RNN+Transformer と表記し図に載せた。

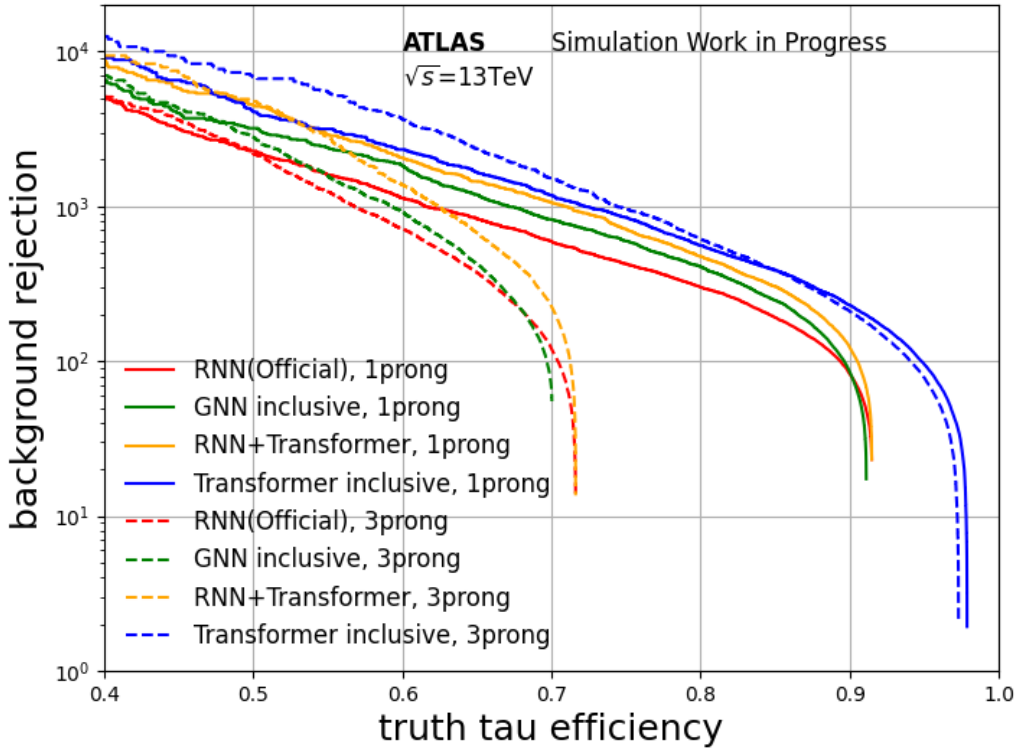


図 7.4: prong 数の判別能力まで含めた背景事象除去能力の比較

表 7.1: 各基準点での背景事象除去能力

モデルの種類	背景事象除去能力		
	0.8	0.6	0.5
RNN, 1prong	300.9	1132.1	2272.7
GNN inclusive, 1prong	408.7	1818.2	3191.5
RNN+Transformer, 1prong	474.7	2054.8	4477.6
Transformer inclusive, 1prong	559.7	2325.6	4109.6
RNN, 3prong	-	717.8	2189.8
GNN inclusive, 3prong	-	917.4	2803.7
RNN+Transformer, 3prong	-	1376.1	4761.9
Transformer inclusive, 3prong	616.0	3658.6	6818.2

全体の傾向としては、背景事象除去能力が低いものから順に RNN、GNN、Transformer となる。Transformer に関しては 3prong では RNN+Transformer よりも Transformer inclusive の方が優れている。1prong では信号事象取得効率が 0.7 より高い領域では Transformer inclusive の方が高いが、それより低い領域ではほとんど差はない。Transformer inclusive で特徴的なのは信号事象取得効率の高さであり最大で 0.97 程度までの値をとり得る。これに対し RNN によるトラック識別や GNN inclusive では 1prong で 0.91、3prong では 0.72 程度の値までしか取ることがない。

図 7.4 において曲線の一端が今までのものと異なり (1,1) に無いのは定義を変えたのが原因である。今までは threshold が 0 をとった場合、信号事象取得効率の分子と分母が同数になり必ず (1,1) をとるようになっていた。しかし定義を変えたことで score の高さだけでなく prong 数の判別結果の正しさも要求するようになり、threshold を 0 にしても分子の数が分母の数を下回るようになった。そのため信号事象取得効率の最大値は prong 数の判別精度を反映している。例えば RNN, 1prong の曲線を見ると最大でも 0.91 程度をとる。これはそもそもトラック識別の RNN が truth 1prong tau の事象を「1prong tau」と判別できる割合が全体の 91%程度しかないことに起因する。

背景事象除去能力の最低値が 1 をとらないのも定義を変えたことが原因である。background のとる score が高かったとしても、prong 数の判別の段階で指定された数以外の prong 数と判定された事象は除去される。そのため今回の定義では prong 数の判別能力が間接的にタウ粒子同定を補助していることになる。例えば RNN の 1prong の場合を見ると、曲線の端点において背景事象除去能力の値は 23.2 をとっている。これは prong 数判別の段階で background が約 1/23 にまで減らされることを意味している。

以上のように背景事象除去能力を見るうえで prong 数の判別精度が関わるようになる。signal と background に対する各モデルでの prong 数の判別結果を図 7.5 に示す。4prong 以上に分類された事象は全て 4prong に入れてヒストグラムを作成している。

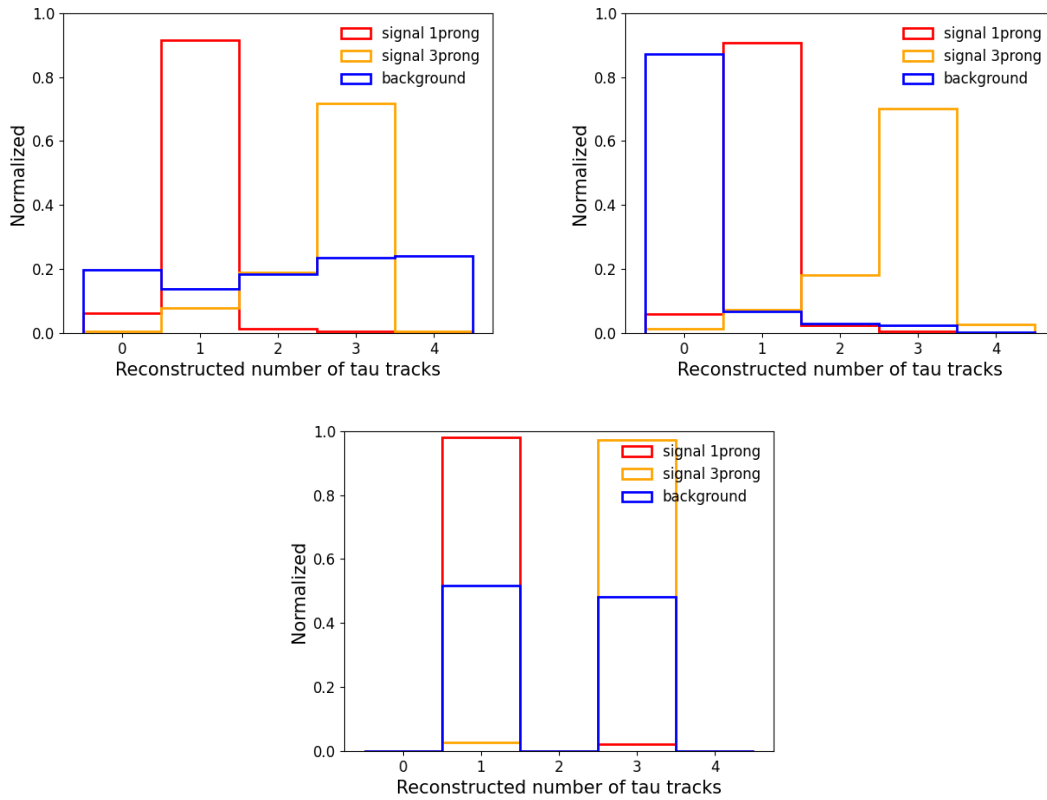


図 7.5: 各モデルでの prong 数判別結果 (左上:RNN、右上:GNN、下:Transformer)

RNN と GNN を見ると signal に対しての prong 数の付け方はどちらも似通っていることがわかる。truth 1prong tau に対しては 1prong と判別されるものが 91%程度あり、次いで多いのが 0prong で 6%程度見られる。truth 3prong tau に対しては正しく prong 数を判別できるものが大きく減り全体の 70%程度になる。次いで多いのが 2prong、その次が 1prong というように実際の prong 数より少なく判別されるものが多い。2prong に分類されたのは全体の 18%程度、1prong が 7%程度である。

background に対しての分布が異なるのは background に対する正解のラベリングが異なるためである。RNN の場合は background でも signal 同様 Tau Track に分類されているトラックの本数が正解のラベルになっているが、GNN では学習に使うコードの技術的な都合上、background の正解のラベルを 0prong と定義している。結果として RNN ではフラットな分布になっているのに対し、GNN では 0prong に強いピークを持っている。タウ粒子同定をする上で重要である 1prong、3prong に分類された事象の割合を見ると signal ではほとんど差が無いのに対し、GNN では 1prong や 3prong に分類された background の数が少ない。prong 数の判別段階で GNN がより background の数を削ることができ背景事象除去能力を上げる要因になると考えられる。

これらに対し Transformer では 1prong または 3prong だけにしか分類しないため、全く異なった分布をとる。signal に対しては 1prong、3prong に関わらず 97%程度を正しい prong 数として判別でき、background に関しては半々に分かれる。

続いて実際の Tau Track の本数分布を示したものが図 7.6 である。

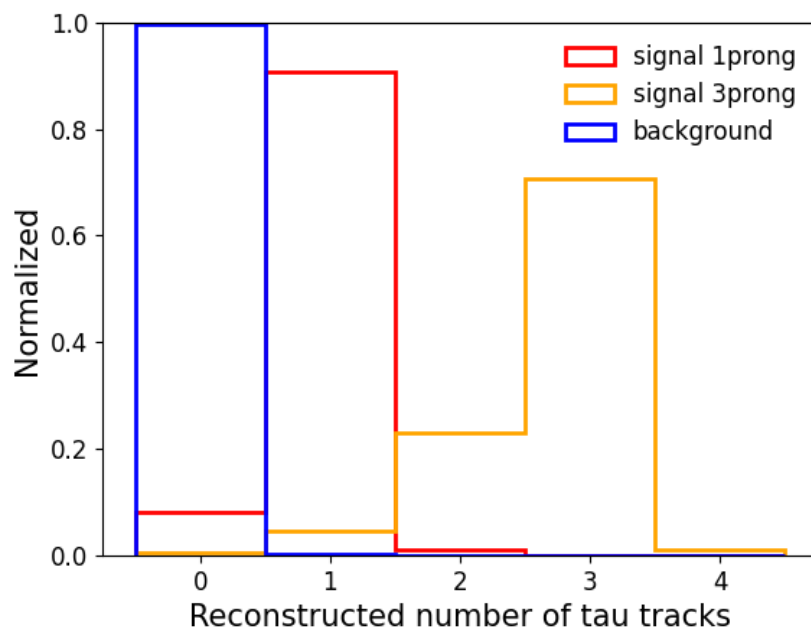


図 7.6: 正解の Tau Track の本数分布

truth 1prong tau では Tau Track を 1 本も含まない事象が全体の 8%程度あり、RNN や GNN のトラック識別による結果と似た分布になることが確認できる。truth 3prong tau でも実際に Tau Track を 3 本含むのは全体の 71%であり、2 本しか含まないものが 23%、1 本のもも 4%程存在する。そのため RNN や GNN の結果で 1prong や 3prong に分類される signal が少ないのは、そもそも本物の Tau Track が少ない事象が存在していることが大きな原因である。これを踏まえると RNN や GNN は正しくトラック識別をしていると言えるが、タウ粒子同定をする上では不都合である。

実際の Tau Track の本数に対して各モデルが prong 数をいくつに判定したかマトリックスにまとめたものを図 7.7 に示す。

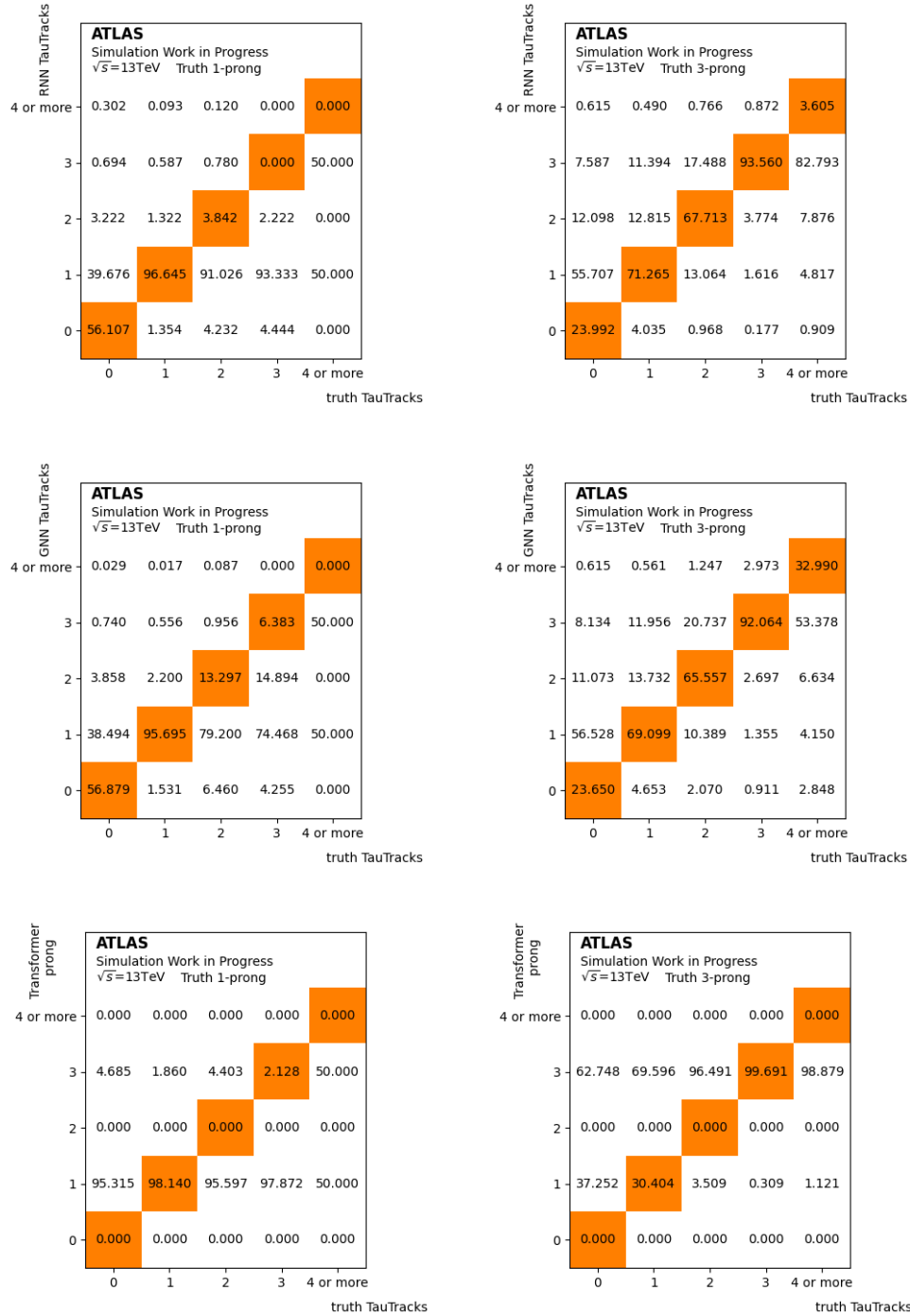


図 7.7: 各モデルにおける正解の Tau Track 数に対する prong 数の判定結果 (上段:RNN、中段:GNN、下段:Transformer)

上段が RNN、中段が GNN、下段が Transformer での結果を表す。また、左側が truth 1prong tau の test data を使用した場合の結果、右側が truth 3prong tau の test data を使用した場合の結果である。それぞれのマトリックスはパーセント表記であり、truth Tau Track の本数ごとに規格化しているため各列の和をとると 100 になるよう変換されているが、実際に足すと丸め誤差の影響で 100 にならない場合がある。横軸が正解の Tau Track の本数であり、縦軸が各モデルで判別した Tau Track の本数である。Transformer は 1prong か 3prong だけにしか分類しないため、0prong や 2prong の

行が全て 0 になっている。

truth 1prong tau を入力にしたもので比較すると、RNN と GNN で truth Tau Track が 1 本の事象に対し Tau Track の本数を 1 本と正しく判断している割合が 95%程度である。しかし truth Tau Track が 0 本の事象は 50%以上を 0 本として分類している。Tau Track が 1 本も含まれていなかった事象であるため分類の仕方としては正しいが、全体の内 8%ある事象の内、半分がこの段階で落とされる。これに対し Transformer では Tau Track が 1 本含まれる事象の 98%、0 本の事象の 95%を 1prong として分類できる。

また、truth 3prong tau を入力した場合を比較すると、Tau Track が 3 本ある場合は 90%以上が 3prong として認識される。Tau Track が 2 本の場合は約 3 分の 2 が 2prong として扱われ、これらの事象はタウ粒子同定の前に排除される。Tau Track が 2 本の事象は truth 3prong tau 全体に対し 23%あるが、その中の 3 分の 2 がタウ粒子同定に進まないため RNN や GNN では信号事象取得効率が大きく落ちる。一方 Transformer では truth Tau Track が 3 本の事象で 99.7%、2 本の場合でも 96.5%を 3prong に分類できる。また truth Tau Track が 1 本の場合、RNN や GNN では 7 割程度を 1prong に分類するが、Transformer では 69.6%を 3prong に分類でき、Tau Track の本数が少ない場合でも RNN などと比べて高い精度で prong 数を判別できることがわかる。

以上のことから Transformer では prong 数を判別する際に高い精度を持っており、truth tau の数がその段階で削られにくいことが背景事象除去能力を上げる要因になっている。Transformer で prong 数の判別精度が高い理由は今後詳しく調べる必要があるが、Transformer ではクラスターやジェットの変数も使用していることが一因として考えられる。RNN や GNN はトラックの情報だけを用いて prong 数を決めているが、Transformer ではトラック以外の変数が有効に働くことで prong 数の判別に役立っている可能性がある。

8 今後の展望

8.1 まとめ

素粒子物理学では精密測定による理論の検証や、理論との矛盾を説明できる新物理探索が課題であり、様々な実験によりこれらが試みられている。LHC-ATLAS 実験では精密測定や新物理探索を行う手法の1つとしてタウ粒子を利用するため、その同定精度が重要となる。ハドロン崩壊したタウ粒子を QCD ジェットと区別するのは難しいため機械学習を導入し、現在は RNN による同定手法が活用されている。

本研究では RNN に代わる深層学習として Transformer を用いたタウ粒子同定手法の開発及び評価を行った。最初は 1prong、3prong のタウ粒子に特化させた学習を行い、既存の RNN や昨年の GNN を使用したモデルよりも高いパフォーマンスを発揮できることを示した。次元数を 64、層数を 4 層または 8 層で試した場合、8 層の方が性能は高くなることが確認できた。

次いで学習用のデータを prong 数で分けずにモデルの学習を行うことで、学習に使用するデータの数を増やし性能の向上を図った。これに伴いトラック識別に使われる変数を追加し、使用するトラックの上限数も 30 本に増やした。結果として特化した学習を行ったときよりも改善が見られ、特に 3prong で背景事象除去能力が向上した。

最後に prong 数の判別までを行う Transformer モデルの学習、評価を行った。既存のトラック識別を用いた手法とは異なり、各事象を 1prong と 3prong だけに分類するようにモデルを作成した。結果として RNN や GNN を使用したモデルよりも高い背景事象除去能力を達成し、特徴として信号事象取得効率の最大値が高くなることが分かった。これは各事象に含まれる本物の Tau Track の本数が prong 数より少ない場合でも、Transformer のモデルであれば prong 数を正しく判別しやすいことに起因している。

8.2 考察

Transformer がなぜ RNN や GNN を使用したモデルよりも高い背景事象除去能力を達成したのか考える。今回使用したモデルで Transformer が RNN や GNN と大きく異なるのがモデルの構造である。例えば RNN ではトラック、クラスターの変数を異なる LSTM 層に入れ演算を行ってから変数の行列を連結、score を計算する。GNN でも同様に主要な計算を先に行った後に変数行列を連結する。これに対し Transformer モデルでは Linear 層で変数の次元を揃えた後、変数行列を連結してから Transformer 層に入れる。そのため Transformer ではトラック、クラスター、ジェットの変数を組み合わせて使用することができる。このようにすべての変数を組み合わせてタウ粒子同定や prong 数の判別を行う点が他のモデルよりも優れた成果を出す理由の可能性がある。例えば prong 数を決める際に Tau Track の情報が無かったとしても、荷電粒子のエネルギーがクラスターの情報として残っていれば prong 数を正しく決められるかもしれない。

また、6 章で書いたようにトラックの本数を増やしたことが精度の向上に大きくつながる可能性がある。前述したようにトラックやクラスターの情報を組み合わせてタウ粒子同定を行えることが Transformer の強みであるため、ここで使える情報を増やすことで更に精度が改善するかもしれない。例えばクラスターの上限数も増やすなどすれば同定能力が上がるのが期待できる。

他にも、学習に使用するデータを増やすことで精度の向上が目指せる。学習用データを prong 数で分けて学習を行うよりも prong 数で分けずに学習させたモデルの方が精度は良いことが確認できている。prong 数で分けなかったことで、トラックの再構成に失敗し Tau Track の本数が少ない事象も含まれているが背景事象除去能力は良くなった。このことから単にデータの数を増やすことが有効であると考えられる。データを簡単に増やす方法としてデータ拡張がある。これは既存のデータを一部欠損させるなどして別のデータを作り出す方法である。例えば各事象のトラックやクラスターを 1

本無くす、今まで横運動量の大きさ順で並んでいたトラックを並び替えるなどが考えられる。prong 数を分けないことで再構成の精度が悪いようなデータを使っても Transformer はよく学習ができたため、データ拡張によるデータのかさましでも性能の向上が見込める。

8.3 今後の展望

Transformer を用いることで現行手法の RNN よりもタウ粒子同定の性能を向上できることを確認したが、どの入力変数や情報が役立っているのかはブラックボックスのままである。score が高い事象と低い事象に分け、それぞれの入力変数分布を確認するなどしてどの変数がタウ粒子同定に役立っているか確認することで今後の研究に役立てることができる。

また、本研究は MC サンプルを用いてモデルの学習、評価を行っているが、実用化するためには実験データにも使用可能か評価を行う必要がある。タウ粒子同定後の実験データと MC サンプルの分布が一致するか確認することでタウ粒子同定が実験データにも機能するか評価する。

加えて、モデルにはまだ改良の余地がある。例えばタウ粒子を解析に使用することを考えるとタウ粒子の電荷を判別する方法も用意する必要がある。本研究で後半に検討した prong 数の判別も行う Transformer ではタウ粒子の電荷まで判別することができないため、そこまでできるようモデルを改良するか電荷の判別を行う別のツールの作成などが有効である。他にも別の深層学習や変数を加えたモデルの検討や、タウ粒子同定、電荷の判別以外の機能を加えたモデルの開発など、モデルの形をさらに改善することが挙げられる。

そしてこれらを通し、精密測定や新物理探索の解析感度を上げることが最終的な目標となる。

9 謝辞

この場を通して、研究を行う上でお世話になった方々への感謝を述べさせていただきます。

まず寄田浩平教授には3年間の研究を通じて、物理だけでなく研究に対する姿勢や考え方までご指導いただきました。私が学部4年生の時は次世代コライダーの検出器の研究、修士からはタウ粒子同定手法の研究を行う機会を頂き、様々な経験を積むことができました。田中雅士主席研究員は研究での関わりこそ少ないものの、研究発表などで意見を頂き説明の仕方や表現などを改善するきっかけを与えていただきました。秘書の坂本さんには研究に伴う出張の手続きなどをしていただいたほか、研究時の差し入れや息抜きの卓球など研究外の部分でもお世話になりました。皆様に心より深く感謝申し上げます。

ATLAS 実験グループの方々にもお世話になりました。寄田研究室のグループメンバーとはミーティングの際に意見を交わし、自分の研究を客観的に見る上で大きな助けになりました。特に、先輩の伊藤さんや秋山さんには研究に対する助言のほか、解析を行う上で必要なソフトウェアの使い方など度々の質問にも答えていただきました。昨年卒業された後藤さんにもタウ粒子の研究を始めたばかりの頃に進め方を教えていただきました。後藤さんの研究内容が私の研究にも関わっており、研究環境の下地を作っていた点でもお世話になりました。また、タウ粒子の研究について話し合う AJ WG1 taus meeting ではタウ粒子や機械学習により詳しい研究者の方々から様々な助言をいただきました。特に ICEPP の森永真央さんにはコーディングなどの技術的な部分から研究方針まで幅広くご意見を頂き大変お世話になりました。皆様、ありがとうございました。

また、同期のおかげで楽しい研究生活を送ることができました。ATLAS グループの荒川君や小川君とは普段の研究や ITk シフトなど様々な場面で助け合いました。研究グループの異なる内海君や石川君とは互いの研究の様子を話すなどして良い息抜きになりました。ありがとうございました。

最後に家族へ感謝を述べたいと思います。修士の卒業まで生活を支えてもらい研究に集中することができました。私が体調を崩し、短い期間ですが入院や通院することになった際もサポートし続けてくれました。ありがとうございました。

A Transformer モデルで次元数と層数を変えた場合の性能比較

一般的に深層学習モデルの性能はパラメータの数が多いほど高くなるが、パラメータ数が増える
と学習するのに十分なデータの数も多くなる。また、モデルが扱う題材や入力変数の数によって必要
なデータ数は変わるため、パラメータ数を変えながら試行錯誤する必要がある。パラメータ数を増や
すには次元数を増やす方法と層数を増やす方法の2パターンが存在する。

本章では5章の内容の前に検証した、モデルの次元数や層数の調整結果について記述する。調整の
際は1prong に特化した Transformer モデルで比較を行った。学習時間の都合から train data のサイ
ズを 480k、validation data のサイズを 120k と小さくしている。signal、background の内訳は 1:1
であり、全て RNN のトラック識別で 1prong と判別されたデータである。

モデルの構造は層数を 8 に固定し次元数を 32、64、128 にしたものと、次元数を 64 に固定し層数
を 4、8、12 にしたものを作成した。

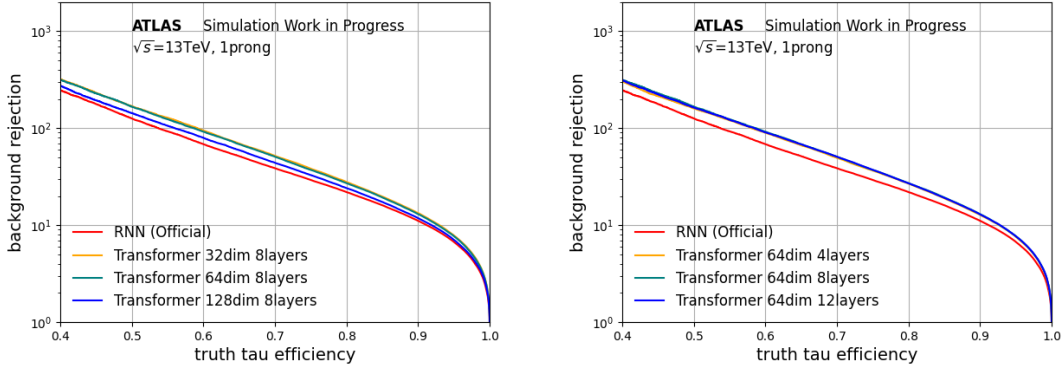


図 A.1: 次元数と層数を変えた場合の比較 (左: 次元数を変動、右: 層数を変動)

またパラメータ数と各基準点での背景事象除去能力を示したものが表 A.1 である。

表 A.1: 次元数、層数を変えた場合の各基準点での背景事象除去能力				
モデルの種類	パラメータ数	背景事象除去能力		
		Loose	Medium	Tight
RNN	-	16.1	29.3	68.5
Transformer, 32dim, 8layers	192097	19.7	38.3	94.4
Transformer, 64dim, 8layers	455553	19.3	37.2	92.0
Transformer, 128dim, 8layers	1203649	17.4	32.9	79.5
Transformer, 64dim, 4layers	255617	19.0	36.5	89.7
Transformer, 64dim, 12layers	655489	19.1	37.2	90.4

8層で固定した場合、次元数は32にしたもので最も背景事象除去能力が高く、次いで64、128と
なる。特に64から128に次元数を増やした場合に大きく下がっていることがわかる。次元数を64に
固定した場合、層数は8にしたものが最も高く、ほとんど差はないが次いで12層、4層と続く。

パラメータが多い方が性能が高いと予想されるにもかかわらず最もパラメータ数が多い128次元、
8層のモデルで性能が低いのは、学習に必要なデータが足りていないためである。層数を8層で固定
し次元数を変えた場合の、学習時の epoch 数に対する loss の推移を図 A.2 に示す。

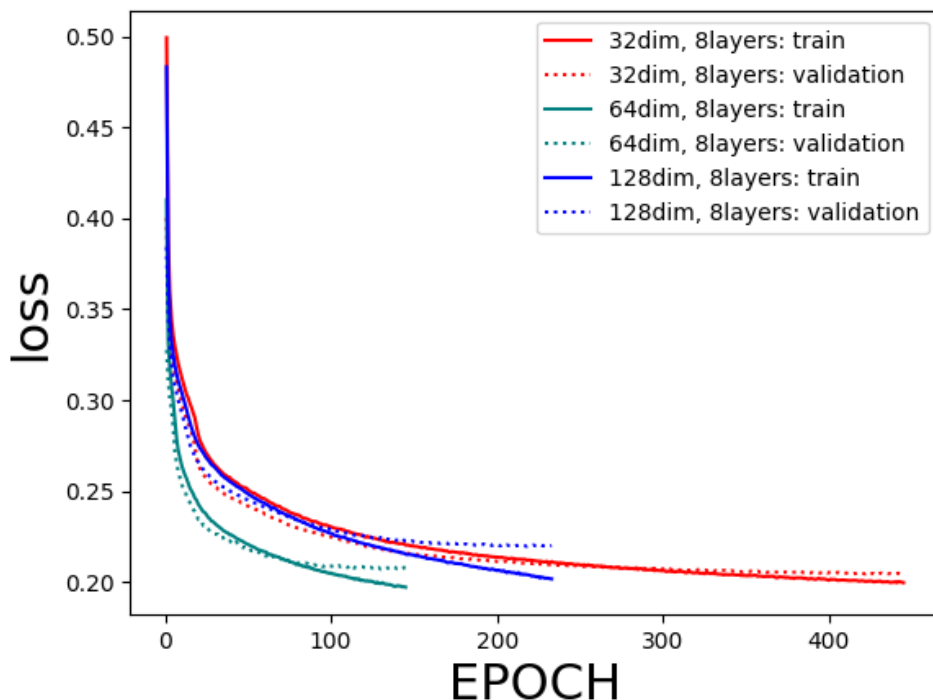


図 A.2: 学習時の epoch 数に対する loss の推移

学習の epoch 数がそれぞれのモデルで異なるのは学習を打ち止める条件によるものである。この学習を行った際は validation data に対する loss が 15 回連続で最小値を更新しなかった場合に学習を打ち切るようになっている。128 次元、8 層のようにパラメータが多い場合、train data に対する loss が順調に下がる一方で、validation data に対する loss が 100 epoch あたりからほとんど横ばいになっており両者で差が開いていることがわかる。これに対し 32 次元、8 層のモデルでは train data と validation data に対する loss が連動して下がり続けている。

このように、パラメータ数が多い場合はモデルが train data に合わせすぎる過学習が起これやすく、結果としてモデルの性能が下がってしまう。そのため大まかなパラメータ数の調整が必要である。また 64 次元、4 層より 64 次元、8 層のほうが性能が良く、32 次元、8 層はそれらよりさらに良い。このことから Transformer では層数を上げつつ次元数を絞るような構造にしたほうが性能が高くなることがうかがえる。このパラメータの調整はモデルの学習時間を短縮する都合から、前述したとおり train data のサイズを 480k にしている。しかし 5 章では 1prong で 3.2M まで train data を増やしている。データサイズが大きくなれば過学習は起きづらくなるため、パラメータ数のある程度上げた方が良いと判断し、5 章では次元数を 64 にしたモデルを使用した。また 12 層にした場合、この試行では 8 層とほとんど結果が変わらない一方で学習時間が長くなったため、層数は 4 層または 8 層で検証することとした。

B 横運動量ごとの背景事象除去能力

5章で学習させたモデルの背景事象除去能力とタウ粒子の横運動量の関係性を調べた。test data をタウ粒子の横運動量の大きさで分け、それぞれで背景事象除去能力を計算することで横運動量に対する依存性を調べる。なお、ここで言うタウ粒子の横運動量とはタウ粒子再構成後のジェットの横運動量であり、truth の値や崩壊前のタウ粒子の持つ運動量ではない。test data の横運動量は基本的に 10GeV ごとで区切っているが、一番端のビンのみ 300GeV 以上の事象も全て含んでいる。また、3prong に関しては高運動量である 150GeV 以上の事象数が少ないため、30GeV 区切りに変えている。

Transformer は次元数を 64、層数を 4 と 8 にしたものを利用して比較を行った。Transformer、GNN のモデルは 1prong では 3.2M、3prong では 1.6M の train data を使って学習させたものである。プロットは score に対する threshold を Loose、Medium、Tight の基準点に合わせ 3 パターンずつ作成した。

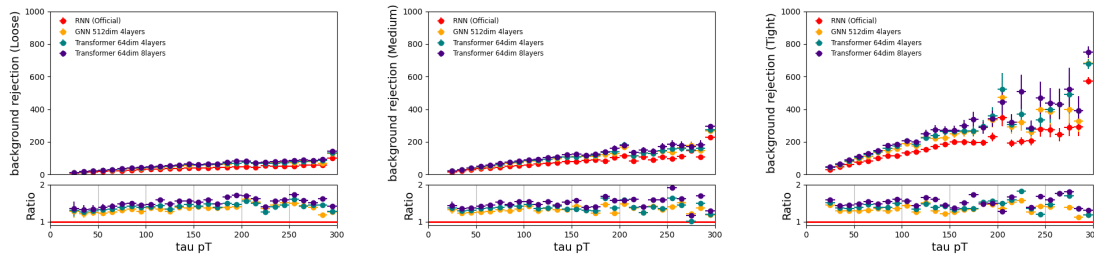


図 B.1: 1prong 事象でのタウ粒子横運動量に対する背景事象除去能力 (左:Loose、中央:Medium、右:Tight)

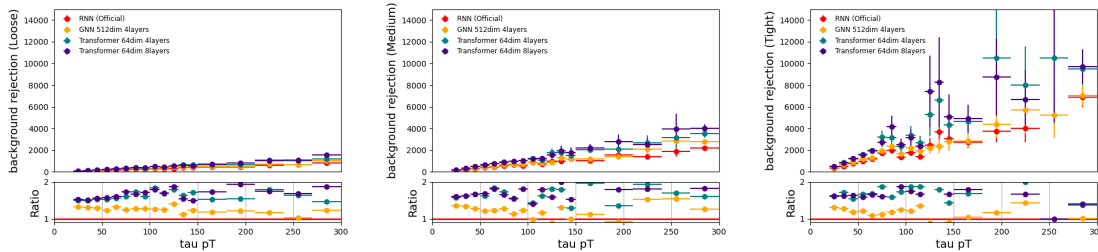


図 B.2: 3prong 事象でのタウ粒子横運動量に対する背景事象除去能力 (左:Loose、中央:Medium、右:Tight)

各図で下側にあるボックスは RNN と各モデルでの背景事象除去能力との比を表している。まず全体を通して言える傾向として、タウ粒子同定の性能はタウ粒子の横運動量が大きくなるほどよくなる。この理由はまだ詳しく調べられていないが、モデルの学習に使用する入力変数の中には横運動量の大きさに依存性を持つものがあるため、signal と background で分布の仕方が変わり区別しやすくなった可能性がある。

RNN との比を確認すると、各横運動量において Transformer では 1.5 付近をとっていることがわかる。そのため Transformer は特定の運動量領域に特化して性能を上げたのではなく、どの運動量でも定常的に RNN から改善している。

C タウ粒子同定と prong 数の判別を行う Transformer での score 分布

prong 数の判別までを行う Transformer モデルで 3 値分類後、規格化前の score 分布を図 C.1、C.2、C.3 に示す。z 軸は規格化してある。

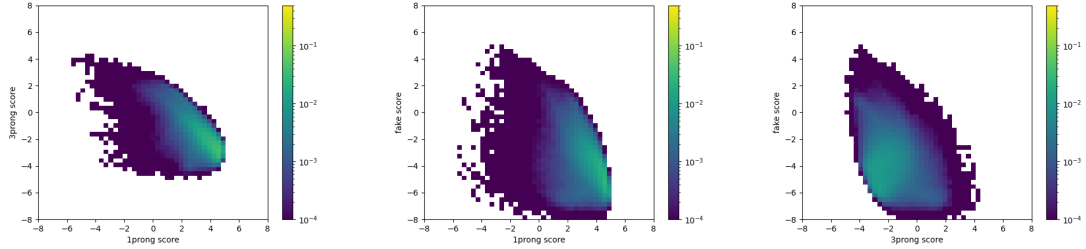


図 C.1: truth 1prong tau に対する各 score の分布

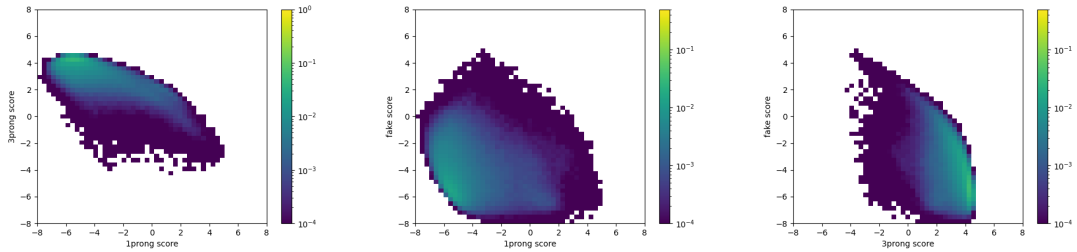


図 C.2: truth 3prong tau に対する各 score の分布

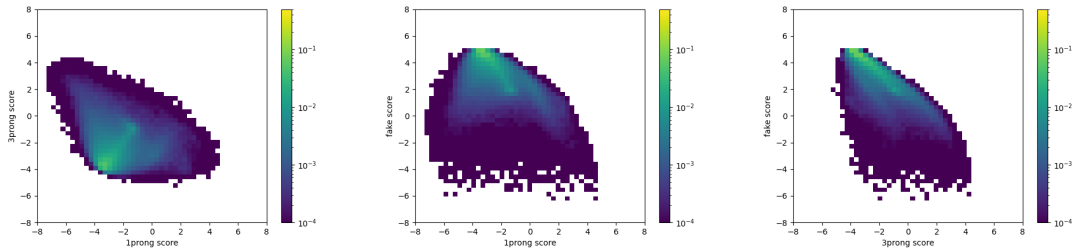


図 C.3: background に対する各 score の分布

1prong と 3prong の score を軸に取っている図において、 $y = x$ より下側に分布しているものは 1prong、上側に分布している事象が 3prong と判別される。しかし truth 1prong tau、truth 3prong tau に対する score では $y = x$ をまたいで尾を引くような分布の仕方をしており、1prong score と 3prong score がある程度近くなる事象が存在することがわかる。そのような事象において規格化後の score が分散してしまうため 1prong と 3prong の score を足す操作を行っている。

また background に対しては 1prong score や 3prong score が -3 付近の値をとるが中でも 1prong score の方が少し値の高い事象が見られる。これは background の中にパイオンが 1 つしか生じないような事象などがあった場合、1prong のタウ粒子と見分けるのが原理的に困難である、といったことが理由として考えられる。

D 2prong も利用した RNN+Transformer と Transformer inclusive の比較

7章では prong 数の判別とタウ粒子同定を行う Transformer モデル (Transformer inclusive と表記) を作成し、今まで RNN のトラック識別で truth 3prong tau かつ 2prong と判別されていたような事象を 3prong に分類することで信号事象取得効率を向上することに成功した。しかし、2prong を使用したいのであればトラック識別の RNN で 2prong と判別された事象を、6章で扱った Transformer モデルで同定する (RNN+Transformer と表記) という方法もとれる。

そこで本章ではこの 2 つの比較を行う。サンプルの信号事象は truth 3prong tau を使用する。RNN+Transformer では 2prong または 3prong と判別された事象が Transformer によるタウ粒子同定に進む。実際に背景事象除去能力を計算したものが図 D.1 である。参考として 2prong と判別された事象を利用していない RNN+Transformer (RNN+Transformer, 3prong Only と表記) と現行手法の RNN の結果も記載した。

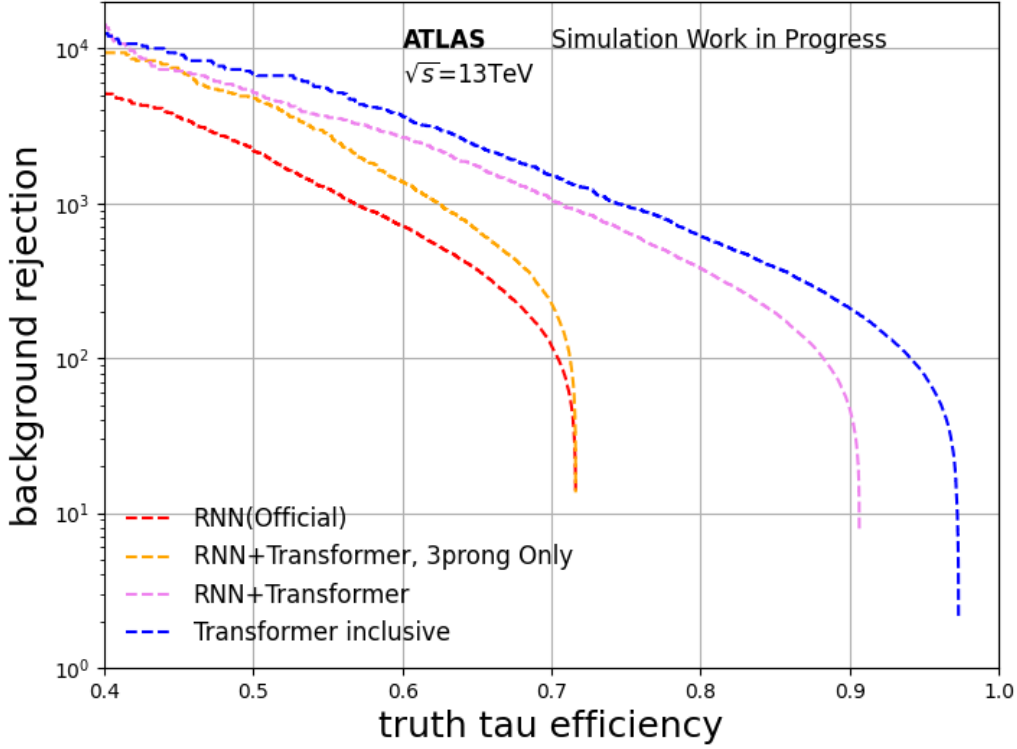


図 D.1: 2prong 事象も利用した RNN+Transformer と Transformer inclusive の比較

また、信号事象取得効率が 0.8、0.6、0.5 の場合の背景事象除去能力を表 D.1 に示す。

表 D.1: 2prong 事象を利用した場合の背景事象除去能力

モデルの種類	背景事象除去能力		
	0.8	0.6	0.5
RNN, 3prong	-	717.8	2189.8
RNN+Transformer, 3prong Only	-	1376.1	4761.9
RNN+Transformer	380.7	2654.9	5172.4
Transformer inclusive	616.0	3658.6	6818.2

2prong と 3prong を使用することでトラック識別用 RNN を使用した場合でも最大で 9 割程度の truth 3prong tau を拾うことができる。これにより現行手法や 3prong と判別した事象のみを使用する Transformer モデルを超えることができる。しかし prong 数の判別とタウ粒子同定を両方行う Transformer モデルの方が prong 数の判別精度、背景事象除去能力の値共に高くなっている。以上のことから、RNN+Transformer で 2prong と判別した事象を使うことで信号事象取得効率を大きく向上させることができるものの、prong 数の判別まで 1 つの Transformer モデルで担った方が良いということがわかる。

E 入力変数を絞った場合の性能比較

今回扱ったモデルでは入力変数が track、cluster、jet の大きく 3 つに分けられるが、それぞれどの程度タウ粒子同定に影響しているか確認するため、入力変数を絞りモデルの学習を行った。

使用したモデルは 7 章のものをベースに、track、cluster、jet いずれかの入力変数を除いたものである。例として jet の変数を使用しなかった場合のモデルを図 E.1 に示す。

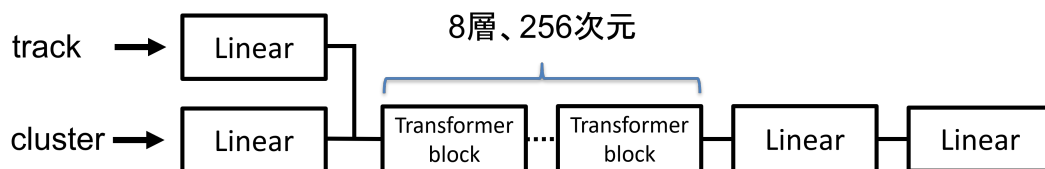


図 E.1: jet の入力変数を使用しない場合の Transformer モデル

単に今まであった jet の入力変数を変換するための Linear 層を無くしたモデルである。このモデルで 3 値分類を行い prong 数の判別とタウ粒子同定を行う。cluster、track の変数を使わない場合も同様に対応する Linear 層を無くしたモデルを使用する。

それぞれの背景事象除去能力を計算した結果が以下である。凡例の内、Transformer trk&cls と表記したものが track、cluster の変数のみを利用し学習したものを指す。trk&jet、cls&jet が track と jet のみ、cluster と jet のみの場合をそれぞれ示している。また、単に Transformer と書いているものは全ての入力変数を使用したものである。

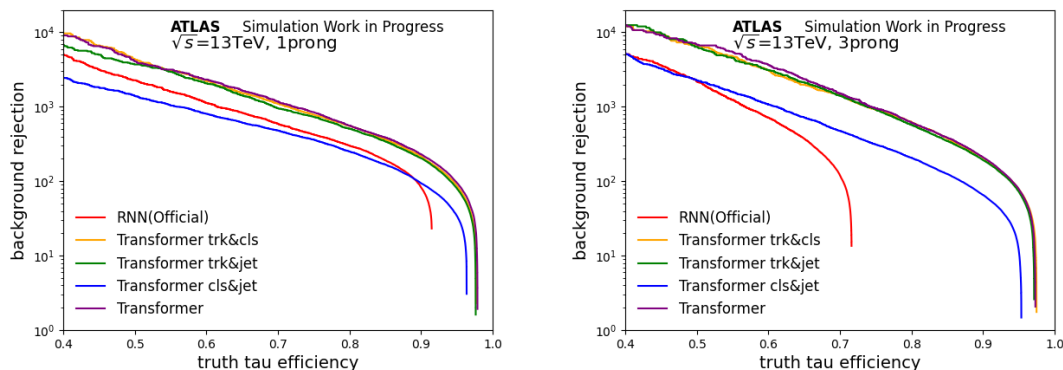


図 E.2: 学習に使用する入力変数を一部除外した場合の比較 (左:1prong、右:3prong)

また、信号事象取得効率が 0.8、0.6、0.5 の場合の背景事象除去能力を表 E.1 に示す。

表 E.1: 各基準点での背景事象除去能力

モデルの種類	背景事象除去能力		
	0.8	0.6	0.5
RNN, 1prong	300.9	1132.1	2272.7
Transformer trk&cls, 1prong	554.5	2205.9	4347.8
Transformer trk&jet, 1prong	508.5	2069.0	3703.7
Transformer cls&jet, 1prong	251.3	806.5	1401.9
Transformer, 1prong	559.7	2325.6	4109.6
RNN, 3prong	-	717.7	2189.8
Transformer trk&cls, 3prong	622.4	3092.8	6521.7
Transformer trk&jet, 3prong	572.5	3092.8	6383.0
Transformer cls&jet, 3prong	206.2	1075.3	2290.1
Transformer, 3prong	616.0	3658.5	6818.2

背景事象除去能力は track の入力変数を使用しなかった場合に大きく下がる。一方で cluster または jet の入力変数を使わなかった場合は、すべての入力変数を使った場合と比べて大きな差は見られなかった。以上のことから track の情報が背景事象の除去に特に有力であることがわかる。Transformer trk&cls、Transformer trk&jet で比較すると前者の方が高い領域が多いことから、track に次いで cluster の変数が役立っていると考えられる。また、track の変数を使用しなくても、1prong で 97% 以上、3prong で 96% 以上、prong 数を正しく判別可能である。現行手法では track の情報のみを使用し prong 数の判別を行っているが、この結果から cluster や jet にも判別に有用な情報が含まれていると推測できる。cluster または jet の入力変数を使用しなかった場合は、すべての入力変数を使用したものと同様に判別の精度が 97% を超える。

track の中にあるどの変数が背景事象を見つけるうえで重要か、など具体的なことはまだわかっていないため、更にモデルを細分化するなどして追加の調査が求められる。

参考文献

- [1] CERN "The Standard Model - ATLAS Physics Cheat Sheet"
<https://cds.cern.ch/record/2851985>
- [2] Particle Data Group "Summary Tables" https://pdg.lbl.gov/2024/tables/contents_tables.html
- [3] CERN "Overall view of the LHC" <https://cds.cern.ch/record/1708847>
- [4] CERN "The HL-LHC project" <https://hilumilhc.web.cern.ch/content/hl-lhc-project>
- [5] The ATLAS Collaboration "The ATLAS Experiment at the CERN Large Hadron Collider"
<https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003/pdf>
- [6] The ATLAS Collaboration "Alignment of the ATLAS Inner Detector in Run-2"
<https://link.springer.com/article/10.1140/epjc/s10052-020-08700-6>
- [7] CERN "Calorimeter"
<https://atlas.cern/Discover/Detector/Calorimeter>
- [8] The ATLAS Collaboration "The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3"
<https://arxiv.org/pdf/2305.16623>
- [9] The ATLAS Collaboration "Operation of the ATLAS Trigger System in Run 2"
<https://arxiv.org/pdf/2007.12539>
- [10] The ATLAS Collaboration "Performance of the ATLAS muon triggers in Run 2"
<https://arxiv.org/pdf/2004.13447>
- [11] The ATLAS Collaboration "Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV"
<https://cds.cern.ch/record/2816332/files/ATLAS-CONF-2022-050.pdf>
- [12] The ATLAS Collaboration "Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV"
<https://arxiv.org/pdf/2002.12223>
- [13] The ATLAS Collaboration "Search for direct stau production in events with two hadronic τ -leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector"
<https://arxiv.org/pdf/1911.06660>
- [14] The ATLAS Collaboration "Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data"
<https://cds.cern.ch/record/2827111/files/ATL-PHYS-PUB-2022-044.pdf>
- [15] Matteo Cacciari, Gavin P. Salam, Gregory Soyez "The anti - k_t jet clustering algorithm"
<https://iopscience.iop.org/article/10.1088/1126-6708/2008/04/063>
- [16] The ATLAS Collaboration "Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV"
<https://arxiv.org/pdf/1412.7086>

- [17] The ATLAS collaboration, "Performance of jet substructure techniques for large- R jets in proton-proton collisions at $\sqrt{s}=7$ TeV using the ATLAS detector"
[https://link.springer.com/article/10.1007/JHEP09\(2013\)076](https://link.springer.com/article/10.1007/JHEP09(2013)076)
- [18] The ATLAS Collaboration "Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment"
<https://cds.cern.ch/record/2688062/files/ATL-PHYS-PUB-2019-033.pdf>
- [19] The Linux Foundation "PyTorch" <https://pytorch.org/>
- [20] The ATLAS Collaboration "Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC"
<https://cds.cern.ch/record/2064383/files/ATL-PHYS-PUB-2015-045.pdf>
- [21] 後藤紘平 "LHC-ATLAS 実験におけるグラフニューラルネットワークを用いたタウ粒子同定手法の開発"
http://www.kylab.sci.waseda.ac.jp/PUB/PAPER/MasterThesis/2023_KoheiGoto.pdf
- [22] The ATLAS Collaboration "Graph Neural Network Jet Flavour Tagging with the ATLAS Detector"
<https://cds.cern.ch/record/2811135/files/ATL-PHYS-PUB-2022-027.pdf>
- [23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin "Attention Is All You Need"
<https://arxiv.org/pdf/1706.03762>
- [24] PyTorch Contributors "ReLU — PyTorch 2.5 documentation"
<https://pytorch.org/docs/stable/generated/torch.nn.ReLU.html>
- [25] PyTorch Contributors "TransformerEncoder"
<https://pytorch.org/docs/stable/generated/torch.nn.TransformerEncoder.html>
- [26] William L. Hamilton, Rex Ying, Jure Leskovec "Inductive Representation Learning on Large Graphs"
<https://arxiv.org/pdf/1706.02216>
- [27] PyTorch Geometric "torch_geometric.nn.conv.SAGEConv"
https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.conv.SAGEConv.html