

修士論文

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

早稲田大学院
先進理工学部研究科 物理学及応用物理学専攻
寄田研究室

5322A030 後藤紘平

2024年2月29日

概要

素粒子物理学における標準模型は多くの実験結果を説明できる一方で、暗黒物質の存在や階層性問題などの解決できない問題もある。これらの問題を解決するために様々な理論が提唱され、様々な実験が世界各地で行われていが、未だ標準模型を超える物理の発見には至っていない。

スイスの欧州原子核研究機構 (CERN) では陽子・陽子衝突型加速器 (LHC) を用いて実験が行われており、現在重心系エネルギー $\sqrt{s} = 13.6$ TeV で稼働している。LHC の 4 つの衝突点の一つを用いた実験が ATLAS 実験であり、標準模型の精密測定や新粒子探索を行っている。

LHC-ATLAS 実験では衝突点に設置している ATLAS 検出器で陽子同士の衝突で発生する粒子の検出を行っており、本研究ではとくにタウ粒子について扱う。ハドロニック崩壊するタウ粒子はジェットとして再構成され、QCD ジェットとの区別が難しいため機械学習を用いた同定が行われている。タウ粒子を用いた解析はヒッグス粒子の精密測定、重いヒッグス粒子の探索、スタウ探索など様々ある。ヒッグス粒子の精密測定は標準模型の検証であり理論とのずれが観測できれば、新物理の発見につながる。ヒッグス粒子の崩壊過程は様々あるが、タウ粒子対への崩壊は分岐比が大きいため重要なチャンネルとなっている。重いヒッグス粒子、スタウは標準模型を超える物理によって存在が示唆される粒子であり、終状態にタウ粒子を含む解析が行われている。しかし未だ発見されておらず様々な実験が質量に制限を与えている。これらの解析の感度を上げるために同定の性能向上が必要である。

同定はカットベースな手法から BDT(決定木)、RNN(Recurrent Neural Network) と、より発展した手法を用いてきた。本研究では新たに GNN(グラフニューラルネットワーク) を用いた手法を開発し、その性能を評価した。

タウ粒子同定には再構成されたタウジェットと付随するトラックとクラスターの情報を用いる。グラフニューラルネットワークは頂点と辺で表されるグラフデータを得意とする深層学習であり、複数存在するトラックやクラスターをグラフに割り当てている。信号事象として $\gamma^* \rightarrow \tau\tau$ 、背景事象として Dijet の MC サンプルを使用し、タウ粒子の崩壊によって発生する荷電粒子のトラック数 (prong) は 1 本と 3 本に分けて学習を行った。性能を最大化するために様々なパラメータの調整を行ったが、GNN の層数と次元の調整について説明する。一般的に機械学習におけるモデルのパラメータ数が大きいほど性能は向上し、そのパラメータ数はデータサンプル数に制限される。層数、次元を増やすとパラメータ数は大きくなるので、どちらを増やすのが効果的か調べたところ、層数を増やす方が効果的であることが分かった。これらのパラメータ調整の結果を用いて、1prong は 3.2M、3prong は 1.6M のデータサンプル数で学習させたところ、現行手法よりも高い背景事象除去能力となった。学習に用いるデータサンプル数をより増やすことで、より性能が向上することが期待される。

さらに、より同定の性能を向上させるためにトラック識別を含めたタウ粒子同定という新しい学習モデルを検討した。トラック識別は同定の前に行われており、そこで prong 数などを定めている。これを同時に行う (トラック識別とタウ粒子同定を包括的に行う) ことで性能が向上するかを調べた。トラック識別の性能の違いを考慮した現行手法との比較において著しい性能向上は見られなかった。データサンプル数を増やす、パラメータ調整など今後の改善が必要である。

目次

1	序論	6
1.1	標準理論	6
1.2	タウ粒子を用いた解析	7
1.2.1	ヒッグス粒子の精密測定	7
1.2.2	重いヒッグス粒子探索	10
1.2.3	スタウ探索	11
2	LHC-ATLAS 実験	13
2.1	LHC	13
2.2	ATLAS 実験	13
2.2.1	ATLAS 検出器の座標系	14
2.2.2	内部飛跡検出器	14
2.2.3	カロリメータ	16
2.2.4	ミューオン検出器	17
2.3	トリガーシステム	18
3	タウ粒子	19
3.1	性質	19
3.2	再構成	19
3.3	トラック識別	21
3.4	同定	23
4	グラフニューラルネットワークを用いたタウ粒子同定	26
4.1	グラフニューラルネットワーク	26
4.2	学習・評価方法	28
4.3	パラメータ調整	33
4.4	現行手法との比較	35
5	トラック識別も含めたタウ粒子同定	37
5.1	学習方法	37
5.2	トラック識別の性能評価	39
5.3	同定の性能評価	43
6	結論と今後の展望	44
7	謝辞	46
A	学習におけるパラメータ調整	47
B	転移学習	48

図目次

1.1	標準理論の素粒子	6
1.2	標準模型におけるヒッグス粒子の崩壊分岐比 [2]	8
1.3	ヒッグス粒子との結合の強さを素粒子の質量の関数として表したもの。青線は標準模型の予想直線 [5]	8
1.4	ヒッグス粒子対生成ダイアグラム	9
1.5	ヒッグス粒子の自己結合定数測定の結果 [6]	9
1.6	重いヒッグス粒子の生成ダイアグラム	10
1.7	タウ粒子を用いた重いヒッグス粒子探索結果 [7]	11
1.8	スタウ対生成の崩壊過程	12
1.9	スタウの探索結果 [8]	12
2.1	LHC の概略図 [9]	13
2.2	ATLAS 検出器の構造 [10]	14
2.3	内部飛跡検出器の構造 [10]	15
2.4	内部飛跡検出器の断面図 [10]	16
2.5	カロリメータの構造 [10]	17
2.6	ミュオン検出器の構造 [10]	17
2.7	ATLAS トリガーシステムの概略図 [10]	18
3.1	生成点再構成効率 [15]	20
3.2	信号事象のトラック識別効率 (左は Truth 1-prong, 右は Truth 3-prong)[15]	22
3.3	Truth 1-prong, Truth 3-prong それぞれにおけるタウトラックの本数の割合 [15]	23
3.4	タウ粒子同定性能 (背景事象取得効率の逆数を信号事象の取得効率の関数であらわしたもの)[15]	25
4.1	グラフの例	26
4.2	GNN で解決できる問題	26
4.3	GNN の学習の流れ [16]	27
4.4	GNN のモデルの種類 [16]	27
4.5	GraphSAGE の演算の流れ [17]	28
4.6	信号事象と背景事象の入力変数分布の違いの例	29
4.7	採用した GNN の学習モデル	29
4.8	入力変数のグラフ化のイメージ図	30
4.9	線形層のイメージ図	30
4.10	学習中の損失の推移	31
4.11	横運動量 p_T 分布	32
4.12	GNNscore 分布の例	32
4.13	タウ粒子同定の性能評価の例	32
4.14	層数と次元を増やしたときの性能変化 (1prong)	33
4.15	データサンプル数を増やした時のモデルのパラメータ数とそのときの背景事象除去能力 (Medium) の変化	34
4.16	現行手法と GNN を用いた手法のタウ粒子同定性能比較	35
4.17	横運動量領域ごとの背景事象除去能力の比較	36

5.1	新たな学習方法の検討	37
5.2	トラック識別も含めたタウ粒子同定の学習モデル	38
5.3	現行手法の信号事象のトラック識別効率 (左: Truth 1-prong, 右: Truth 3-prong) . . .	39
5.4	GNN を用いた手法の信号事象のトラック識別効率 (左: Truth 1-prong, 右: Truth 3-prong)	40
5.5	現行手法と GNN を用いた手法における信号事象の prong 数 (Truth 1-prong)	40
5.6	現行手法と GNN を用いた手法における信号事象の prong 数 (Truth 3-prong)	41
5.7	現行手法と GNN を用いた手法における背景事象の prong 数	41
5.8	GNN を用いた手法によって識別された prong 数 (現行手法で 3prong と識別されたイベントのうち)	42
5.9	0prong と 3prong と識別されたものの RNNscore 分布	42
5.10	現行手法との性能比較 (左: 1prong, 右: 3prong)	43
6.1	横運動量領域における信号事象 truth 3-prong の prong 数	45
付録.1	Dropout のイメージ図 [18]	47
付録.2	Dropout の有無のよる学習の推移の違い (赤: Dropout あり、緑: Dropout なし) . . .	47
付録.3	high の割合を変えた時の性能変化	49

表目次

3.1	タウ粒子の崩壊分岐比	19
3.2	トラック識別で使用している入力変数	22
3.3	タウ粒子同定で使用している入力変数	24
3.4	タウ粒子同定の信号取得効率の基準点	25
4.1	サンプルのカット条件	28
4.2	基準点 Medium における層数と次元を増やしたときの性能変化	33
4.3	基準点 Medium における横運動量領域ごとの背景事象除去能力の比較	35
5.1	サンプルのカット条件の変更点	37

1 序論

1.1 標準理論

素粒子とは物質を構成する最小単位の粒子とそれらの相互作用を媒介する粒子のことを指す。現在は標準理論で説明される 17 の素粒子の存在が確認されている。素粒子の一覧を図 1.1 に示す。



図 1.1: 標準理論の素粒子

自然界において電磁相互作用、弱い相互作用、強い相互作用、重力相互作用の 4 つが存在すると考えられており、標準理論は重力相互作用を除いた 3 つの相互作用を説明するものである。17 種類の素粒子は物質を構成するフェルミオン、相互作用を媒介するゲージボソン、質量の起源とされるヒッグス粒子に分けられる。それぞれについて説明する。

1. フェルミオン

スピンの大きさが $1/2$ の粒子であり、クォークとレプトンに分けられ、それぞれ第 1 世代から第三世代の 3 つのセットが存在している。

(a) クォーク

カラー荷をもち、強い相互作用をする素粒子であり、図 1 に示すように 6 種類存在する。上段のアップクォーク、チャームクォーク、トップクォークは $+2/3$ の電荷をもち、下段のダウンクォーク、ストレンジクォーク、ボトムクォークは $-1/3$ の電荷をもつ。クォーク単体では

存在せず、クォークが複数結合したハドロンとして存在する。特にクォークが2つ結合したものをメソン、3つ結合したものをバリオンと呼ぶ。

(b) レプトン

カラー荷をもたず、弱い相互作用をする素粒子であり、図1に示すように6種類存在する。下段の電子、ミューオン、タウとそれに対応する上段の電子ニュートリノ、ミューニュートリノ、タウニュートリノがあり、前者は電荷が-1で後者は電荷が0である。ニュートリノは電荷が0かつカラー荷をもたず、弱い相互作用しか起こさないため、検出するのが難しい粒子である。

2. ゲージボソン

スピンの大きさが1の粒子であり、図1に示すようにグルーオン、Wボソン、Zボソン、光子の4つがある。グルーオンは強い相互作用を媒介する粒子であり、質量は0である。カラー荷を2つもち、8種類存在する。Wボソン、Zボソンは弱い相互作用を媒介する粒子であり、Wボソンは約80GeV、Zボソンは約91GeVの質量である。光子は電磁相互作用を媒介する粒子であり、質量は0である。

3. ヒッグス粒子

ヒッグス機構により存在が予言され、2012年に発見された。スピンの大きさが0の粒子であり、質量は約125GeVである。

1.2 タウ粒子を用いた解析

本論文ではタウ粒子について取り扱うので、LHC-ATLAS実験におけるタウ粒子を用いた解析の一部について説明する。

1.2.1 ヒッグス粒子の精密測定

2012年に新粒子がATLAS・CMS実験の両実験で3つの崩壊モードで同じ質量領域に見つかった。これらの解析を統合すると、新粒子の質量は $126.0 \pm 0.4(\text{stat}) \pm 0.4(\text{syst})\text{GeV}$ で有意度は 5.1σ であった。[1]その後性質測定が進み、スピンの大きさが0であること、生成断面積や崩壊分岐比が標準理論で予想されるヒッグス粒子のものとほぼ一致していることが判明した。このヒッグス粒子の発見は標準理論の完成を意味した。

標準理論は様々な実験結果において高い精度で検証されている。しかし、標準理論は階層性問題、ニュートリノの軽さ、暗黒物質の存在などを説明することができない。このことから標準理論を超える新物理の存在が示唆されている。標準理論とのずれを測定することが新物理探索において重要であり、とくにヒッグス粒子の湯川結合定数と自己結合定数の精密測定が重要である。

湯川結合定数

標準模型においてフェルミオンとヒッグス粒子の湯川結合の大きさはフェルミオンの質量に比例する。湯川結合定数を測定することでフェルミオンの質量獲得の描像を検証することができる。

標準模型におけるヒッグス粒子の崩壊分岐比を図1.2に示す。 t クォークは重すぎてヒッグス粒子の崩壊では生成されず、 b クォーク対への崩壊が最も高い分岐比をもつ。これまでにフェルミオンの中でも質量の重い第三世代の素粒子(τ レプトン, b クォーク, t クォーク)の湯川結合は発見された。 t クォークの湯川結合の測定には $t\bar{t}H$ 生成過程を利用した。崩壊分岐比の大きい $H \rightarrow b\bar{b}$ を用いる解析が期待されていたが、 $t\bar{t}$ に付随して b クォークが生成される過程の理論的な不定性が大きいため、 $H \rightarrow \gamma\gamma$ 、 $H \rightarrow WW$ 、

$H \rightarrow \tau\tau$ を用いた解析がより高い感度となった。第二世代のフェルミオンとの湯川結合の検証はミュオンで行われている。第二世代のフェルミオンの中で一番湯川結合が大きいのは c クォークであるとあると考えられるが、強い相互作用由来の背景事象の影響で $H \rightarrow c\bar{c}$ 事象を発見することは現時点では難しい。 $H \rightarrow \mu\mu$ 事象の探索では $Z \rightarrow \mu\mu$ 事象が主な背景事象である。終状態の光子のエネルギーを考慮することで $\mu\mu$ の不変質量の分解能を向上させ、ATLAS 実験の Run2 解析 [3] で有意度 2.0σ 、CMS 実験 [4] で有意度 3.0σ を得た。Run3 解析で ATLAS 実験と CMS 実験の結果を統合することで 5.0σ の有意度が得られる可能性がある。

これまでに測定された湯川結合定数の結果をまとめたものを図 1.3 に示す。 κ は結合定数を標準模型の予想値で規格化したものである。 $\kappa_Z, \kappa_W, \kappa_t$ は 10% 以下、 κ_b, κ_τ は 15% 程度、 κ_μ は 30% 程度の精度で標準模型と一致している。今後は第三世代の湯川結合定数の測定の精度を上げること、第二世代のフェルミオンとの結合の検証が目標である。

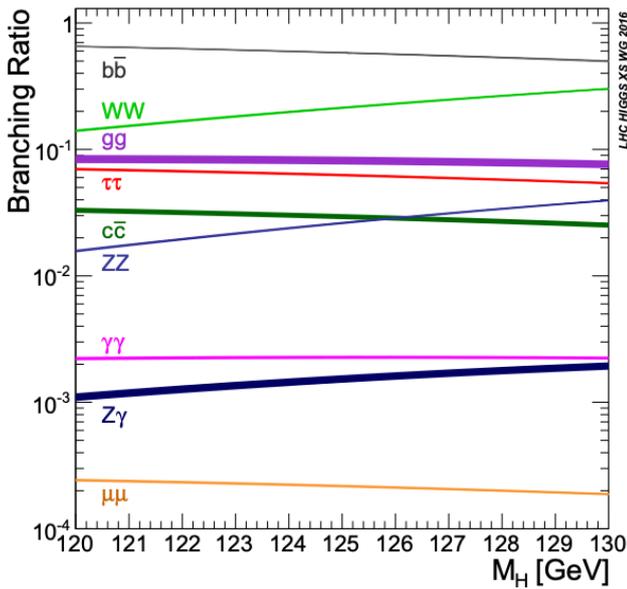


図 1.2: 標準模型におけるヒッグス粒子の崩壊分岐比 [2]

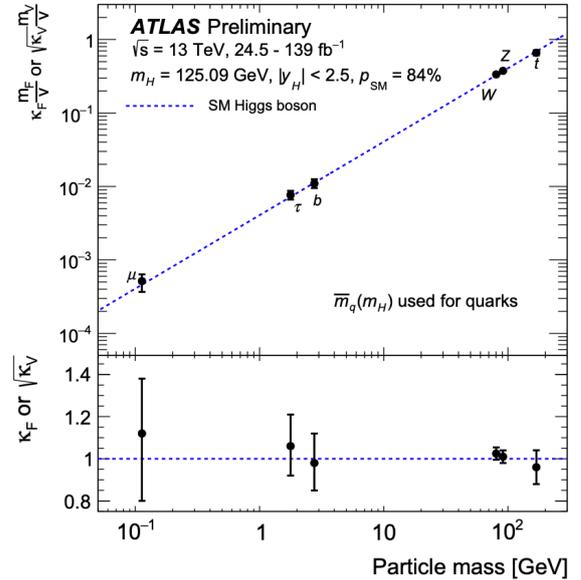


図 1.3: ヒッグス粒子との結合の強さを素粒子の質量の関数として表したもの。青線は標準模型の予想直線 [5]

自己結合定数

フェルミオンとボソンの相互作用はゲージ理論によって記述される。ゲージ不変であると力を媒介する粒子は質量が 0 でなければいけない。初期宇宙ではすべての素粒子の質量が 0 であったとされているが、現在の宇宙では W ボソンや Z ボソンは質量を持っている。このことを電弱相互作用におけるゲージ対称性の破れと呼ぶ。

この問題の解決のために導入されたのがヒッグス機構である。スピン 0 のスカラー場 ϕ のポテンシャル

$$V(\phi) = \lambda(|\phi|^2 + \frac{\mu^2}{2\lambda}) \quad (1.1)$$

を導入する。 λ はヒッグス粒子の自己結合定数で、 μ^2 は宇宙誕生直後の高温では 0 より大きく、低温になると 0 より小さくなるパラメータである。 $\mu^2 < 0$ の場合 V は有限の ϕ で最小値をとり、その最小値を与える ϕ を真空期待値 $v = \sqrt{-\mu^2/\lambda}$ と呼ぶ。宇宙初期の高温では ϕ の期待値は 0 だが、現在の宇宙では $\langle \phi \rangle = v/\sqrt{2}$ となるので、 W ボソンと Z ボソンは質量を獲得することができる。

式 (1.1) から分かるようにヒッグス粒子の自己結合定数の測定はヒッグス機構の直接的な検証のために重要である。自己結合定数の測定はヒッグス粒子対生成事象を用い、図 1.4 のダイアグラムが支配的である。ヒッグス粒子対生成事象の断面積はヒッグス粒子単体生成の断面積の千分の一以下と小さく、探索は非常に困難である。 $HH \rightarrow b\bar{b}\tau^+\tau^-$ 、 $HH \rightarrow b\bar{b}b\bar{b}$ 、 $HH \rightarrow b\bar{b}\gamma\gamma$ などの崩壊モードを用いて探索が行われている。図 1.5 に Run2 の $126\text{--}139\text{ fb}^{-1}$ のデータを用いた解析結果を示す。3つの崩壊モードの解析を統合すると標準模型の約 2 倍の生成断面積まで棄却された。

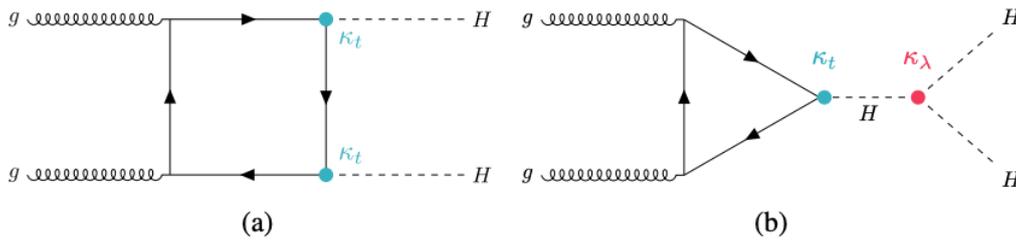


図 1.4: ヒッグス粒子対生成ダイアグラム

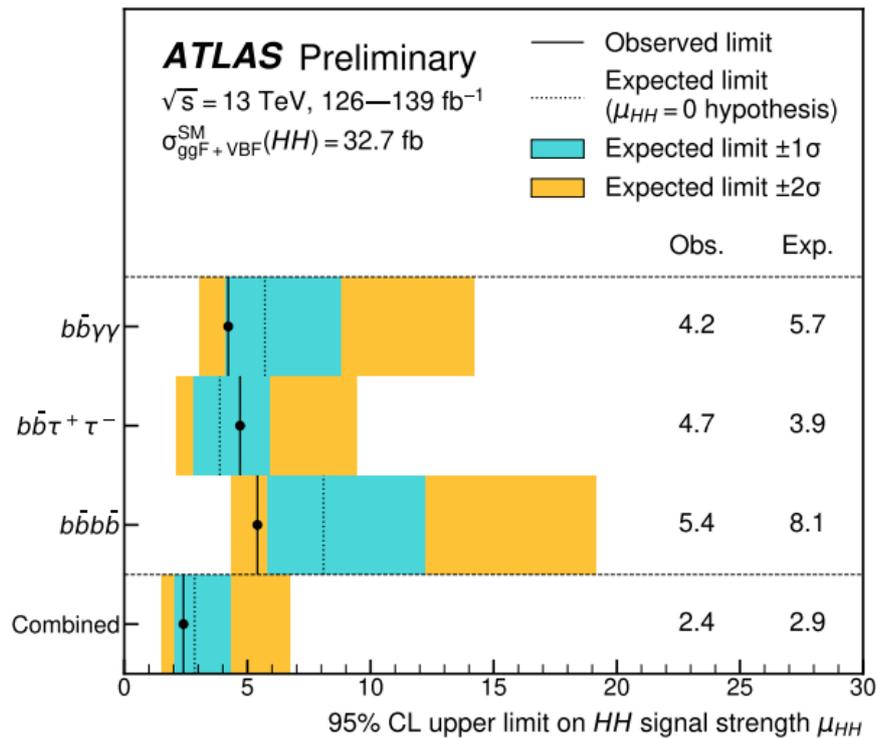


図 1.5: ヒッグス粒子の自己結合定数測定の結果 [6]

湯川結合定数、自己結合定数の測定は標準理論とのずれがあるか否かを検証するために重要である。とくに自己結合定数は多くのモデルによって標準理論の値からずれる可能性が指摘されていて、それは新物理の発見であるとともに、自己結合定数の大きさは物質優勢宇宙の起源の一つである電弱バリオジェネシスにおいて重要な役割を果たす。

1.2.2 重いヒッグス粒子探索

超対称性理論などの標準理論を超えた物理では Two Higgs Doublet Model(2HDM) に代表される拡張ヒッグスセクターの存在が予想されていて、ヒッグス粒子が複数存在する可能性がある。このモデルでは H_u^0 と H_d^0 という2種類のヒッグス粒子が導入され、それぞれの真空期待値 ν_u, ν_d は $\nu = \sqrt{\nu_u^2 + \nu_d^2}$ を満たす。 $(\nu$ は標準理論における真空期待値) 標準理論のヒッグス (h) に加えて、重いスカラーヒッグス (H)、重い擬スカラーヒッグス (A)、荷電ヒッグス (H^\pm) の4つのヒッグス粒子が導入される。2HDM にはいくつかの種類があり、その中の Type II と呼ばれるものは Minimal Supersymmetric Standard Model(MSSM) におけるヒッグスセクターとなっている。MSSM は超対称性を実現する最小の標準模型拡張モデルのことである。MSSM のヒッグスセクターはアップタイプのクォークとの結合は小さく、ダウンタイプのクォークとの結合は大きくなる。また MSSM のヒッグスセクターは重い擬スカラーヒッグス A の質量 m_A と真空期待値の比 $\tan\beta = \nu_u/\nu_d$ のみに依存する。

$\tan\beta$ の大きい領域ではダウンタイプのクォークとの結合が大きくなるため、b クォークを伴う過程の生成断面積は大きくなる。また $\tan\beta >$ が大きくなるとタウ粒子や b クォーク対への崩壊分岐比が大きくなる。このような点からヒッグス粒子が b クォーク、タウ粒子対へ崩壊するモードの探索がこれまでも行われてきた。ATLAS 実験 Run2 の 139 fb^{-1} のデータを用いた解析 [7] では図 1.6 で示す生成過程でヒッグス粒子がタウ粒子対に崩壊するモードを用いて探索が行われた。解析結果を図 1.7 に示す。 $m_A=1.0 \text{ TeV}$ で $\tan\beta > 8$ 、 $m_A=1.5 \text{ TeV}$ で $\tan\beta > 21$ を 95 %信頼区間で棄却している。この解析では $\tau_{lep}\tau_{had}$ チャンネル (レプトン崩壊とハドロン崩壊のタウ粒子それぞれ1つ) と $\tau_{had}\tau_{had}$ チャンネル (ハドロン崩壊のタウ粒子が2つ) を用いられている。後者ではシングルタウトリガーを使用している。その p_T の閾値はデータ取得年度によって異なり、80 GeV, 125 GeV, 160 GeV のものが使われている。図 1.7 の解析結果にあるように m_A の質量が重い領域が未探索領域となっており、この領域では高横運動量タウ粒子が生じる。後の章でも説明するがタウは QCD ジェットとの区別が難しいため、機械学習による同定が行われている。高横運動量領域におけるタウ粒子の同定性能を向上させることで、より m_A の質量が重い領域の探索が可能となる。

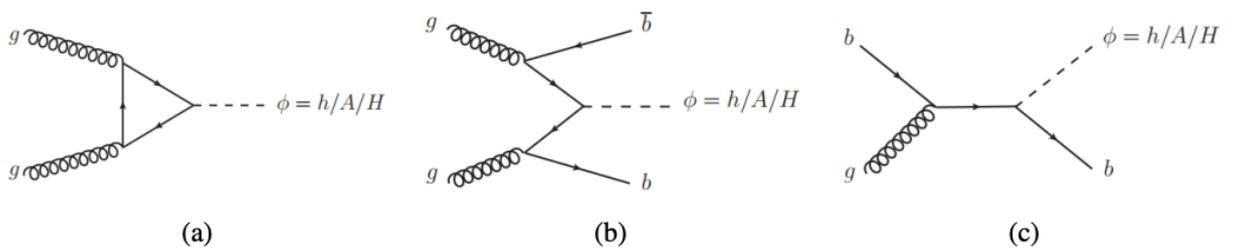


図 1.6: 重いヒッグス粒子の生成ダイアグラム

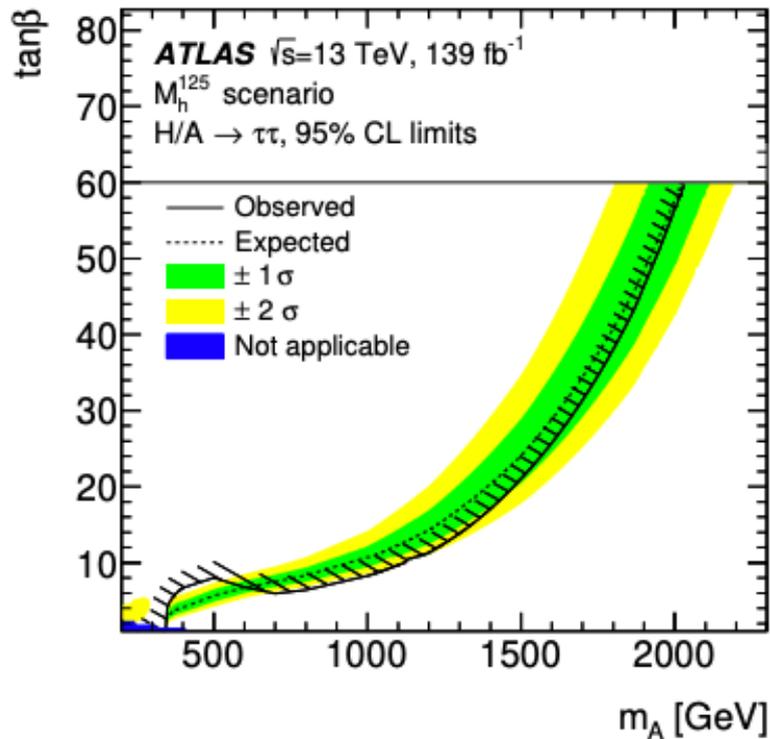


図 1.7: タウ粒子を用いた重いヒッグス粒子探索結果 [7]

1.2.3 スタウ探索

標準理論には階層性問題や暗黒物質の存在を説明できないなどいくつかの問題がある。上記の問題を解決する新物理の有力な候補として超対称性理論 (SUSY) がある。超対称性はボソンとフェルミオンの対称性で、標準模型の素粒子に超対称性パートナーが存在する。それらは標準模型の素粒子とスピン以外の量子数が同じでスピンだけ異なる。フェルミオンのパートナーのスピンは 0、ボソンのパートナーのスピンは 1/2 である。超対称性が成り立っていれば電子のパートナーの質量は電子と同じであるが、そのような粒子は見つかっていない。それは超対称性の破れによって、超対称性粒子は元の標準模型の素粒子よりも重くなると考えられている。

スタウはタウの超対称性パートナーであるが、現時点では見つかっておらず様々な実験でスタウの質量に制限を与えている。ATLAS 実験 Run2 の 139 fb^{-1} のデータを用いた解析 [8] では図 1.8 のダイアグラムを信号事象として探索した。スタウが対生成し、タウとニュートラリーノ ($\tilde{\chi}_1^0$) に崩壊する過程である。ニュートラリーノは超対称性理論で存在が予想されている粒子で、ズィーノ (Z ボソンのパートナー)、フォティーノ (光子のパートナー)、中性ヒグシーノ (中性ヒッグスのパートナー) による混合状態である。ハドロン崩壊するタウ 2 つとタウから発生するニュートリノ、ニュートラリーノによる消失横運動量などを信号領域で要求している。

図 1.9 が解析結果であり、スタウとニュートラリーノの質量に制限を与えている。スタウとニュートラリーノの質量差が小さい領域が未探索となっており、この領域を探索するためには低横運動量タウ粒子を要求する必要がある。しかし ATLAS 実験では膨大な事象が発生しすべての事象を記録することができないためトリガーを設けており、この解析で使用しているトリガーはタウの横運動量に閾値を設けている。QCD ジェットの背景事象を減らし、タウトリガーの閾値を下げることで未探索領域の感度を上げることに繋がる。

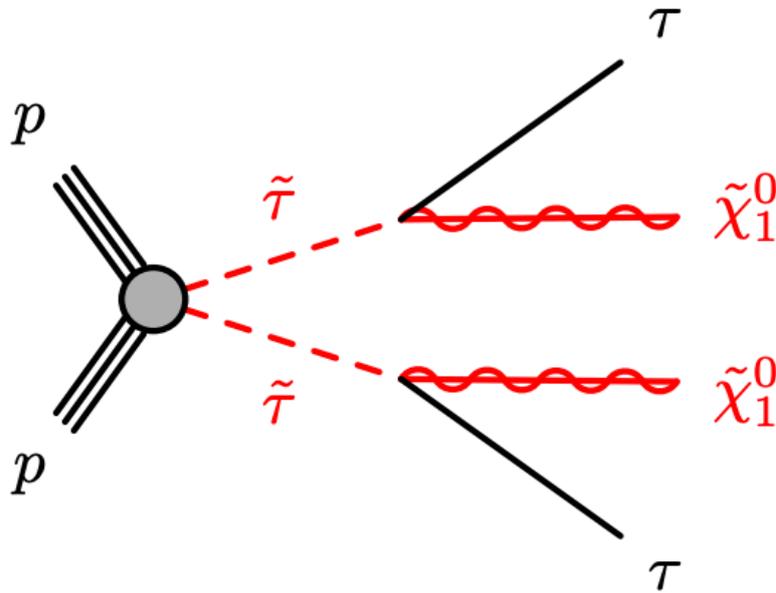


図 1.8: スタウ対生成の崩壊過程

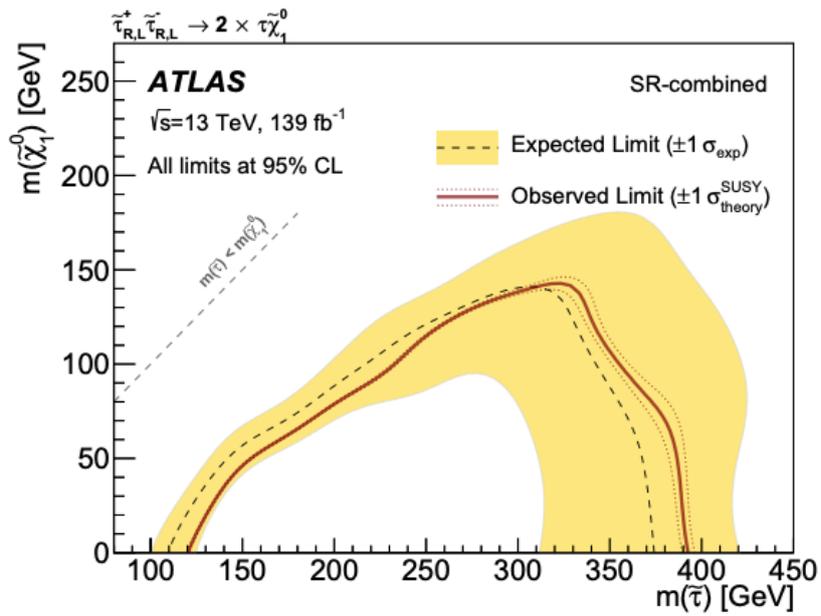


図 1.9: スタウの探索結果 [8]

2 LHC-ATLAS 実験

2.1 LHC

LHC(Large Hadron Collider) は CERN に建設された陽子陽子衝突型加速器である。陽子と陽子を光速に限りなく近い速さまで加速し衝突させ、その衝突によって発生する粒子を検出することで、標準理論の精密測定や標準理論を超えた物理の発見を目指している。LHC はスイスとフランスの国境付近の地下約 100 メートルに建設され、周長は約 27km ある。LHC の概略図を図 2.1 に示す。2011 年から 2012 年にかけて Run1 が重心エネルギー 7TeV で行われ、2015 年から 2018 年にかけて Run2 が重心系エネルギー 13 TeV で行われた。2022 年からは Run3 が重心系エネルギー 13.6 TeV で行われている。また 2029 年頃に HL-LHC(High Luminosity LHC) の開始が計画されている。HL-LHC は陽子と陽子の衝突頻度を上げてデータ量を約 10 倍に増やすことによって、新粒子探索感度の向上や、ヒッグス粒子の精密測定などを目指している。(それに伴い、ATLAS 検出器のアップグレードも予定されており、そのアップグレードのための部品の量産、量産体制の準備を現在日本を含めた ATLAS 実験への参加国が行っている。)

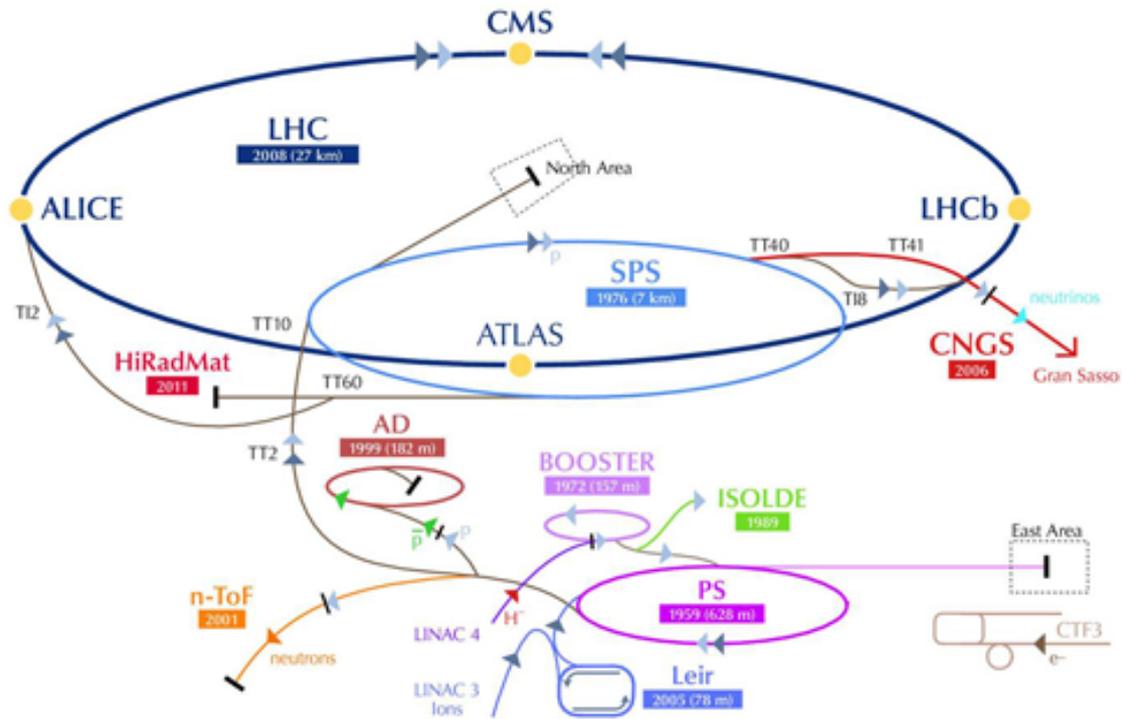


図 2.1: LHC の概略図 [9]

2.2 ATLAS 実験

ATLAS 実験は LHC を用いた実験の 1 つであり、全長 46 メートル、直径 25 メートル、重量 7000 トンの ATLAS 検出器を衝突点に配置している。図 2.2 に ATLAS 検出器の構造を示す。ATLAS 検出器は内側から内部飛跡検出器、電磁カロリメータ、ハドロンカロリメータ、ミューオン検出器といった構成になっている。

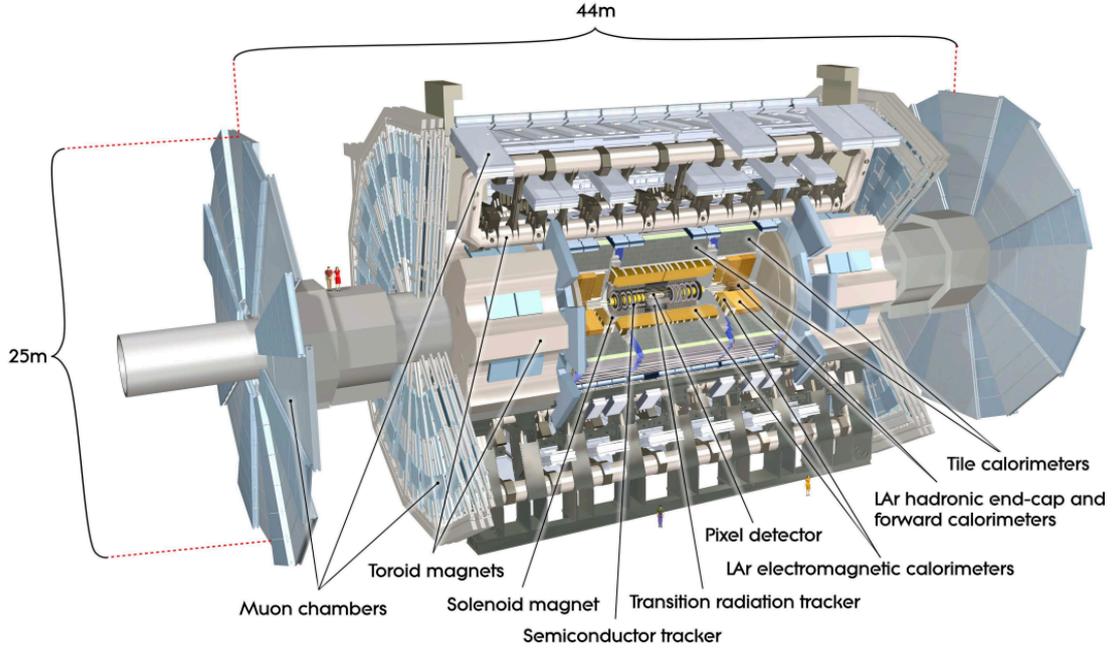


図 2.2: ATLAS 検出器の構造 [10]

2.2.1 ATLAS 検出器の座標系

ATLAS 検出器における座標系とよく使われる変数について説明する。ATLAS 検出器における座標系は原点をビームラインの衝突点の中心にとり、LHC のリングの中心方向を x 軸の正の方向、地上方向を y 軸の正の方向とする右手系である。良く使われる変数として横運動量 p_T 、方位角 ϕ 、擬ラピディティ η と 2 粒子間の距離を表す ΔR などがある。横運動量 p_T は式 (2.1) で定義され、 x - y 平面上で運動量が保存するため良く使われる。

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (2.1)$$

擬ラピディティ η は天頂角 θ を用いて、以下に示す式 (2.2) で定義される。

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (2.2)$$

ΔR は 2 点間の $\Delta\eta$ 、 $\Delta\phi$ を用いて、以下に示す式 (2.3) で定義される。

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (2.3)$$

2.2.2 内部飛跡検出器

内部飛跡検出器は ATLAS 検出器の最内層に位置する。内部飛跡検出器の構造を図 2.3 に、ビーム軸方向を横軸とした構造を図 2.4 に示す。 $|\eta| < 2.5$ の領域で荷電粒子の飛跡を検出し、運動量や衝突点の

位置を測定する。ソレノイド磁石により 2T の磁場がかかっているため、荷電粒子の飛跡の曲率から運動量を計算することができる。内部飛跡検出器は内側から Insertable B-layer (IBL)、pixel 検出器、シリコンストリップ検出器 (SCT)、遷移放射型検出器 (TRT) といった構成になっている。ビーム軸に並行なバレル領域とビーム軸に垂直なエンドキャップ領域に分かれている。

- IBL, pixel

IBL, pixel は半導体検出器で構成され、荷電粒子の通過位置を 2 次元で読み出す。IBL は Run2 から追加された層で、ビーム軸から 33mm の位置にあり、ピクセルのサイズは横方向 $50 \times 250 \mu\text{m}^2$ である。残りの 3 層の pixel 検出器はビーム軸から 50.5、88.5、122.5mm の位置にあり、ピクセルのサイズは $50 \times 400 \mu\text{m}^2$ である。位置分解能は動径方向と方位角方向で $10 \mu\text{m}$ 、z 軸方向に $115 \mu\text{m}$ である。

- SCT

semiconductor tracker (SCT) は pixel 検出器の外側に位置し、バレル領域 4 層、エンドキャップ 9 層からなるストリップ型の半導体検出器である。ストリップの長さは 126mm で $80 \mu\text{m}$ 間隔にならんでいる。ストリップは片方がビーム方向と並行で、もう片方を 40mrad ずらし合わせることで 2 次元での読み出しを行っている。位置分解能は動径方向と方位角方向で $17 \mu\text{m}$ 、z 軸方向に $580 \mu\text{m}$ である。

- TRT

遷移放射検出器 (TRT) は荷電粒子が異なる物質間を移る際に起こる現象である遷移放射を利用する検出器である。内部は $X_e, \text{CO}_2, \text{O}_2$ の混合気体で満たされており、荷電粒子が通過したときに生じる電荷を検出する。直径 4mm のドリフトチューブの積み重ねることによって構成されていて、位置分解能は動径方向と方位角方向で $130 \mu\text{m}$ である。

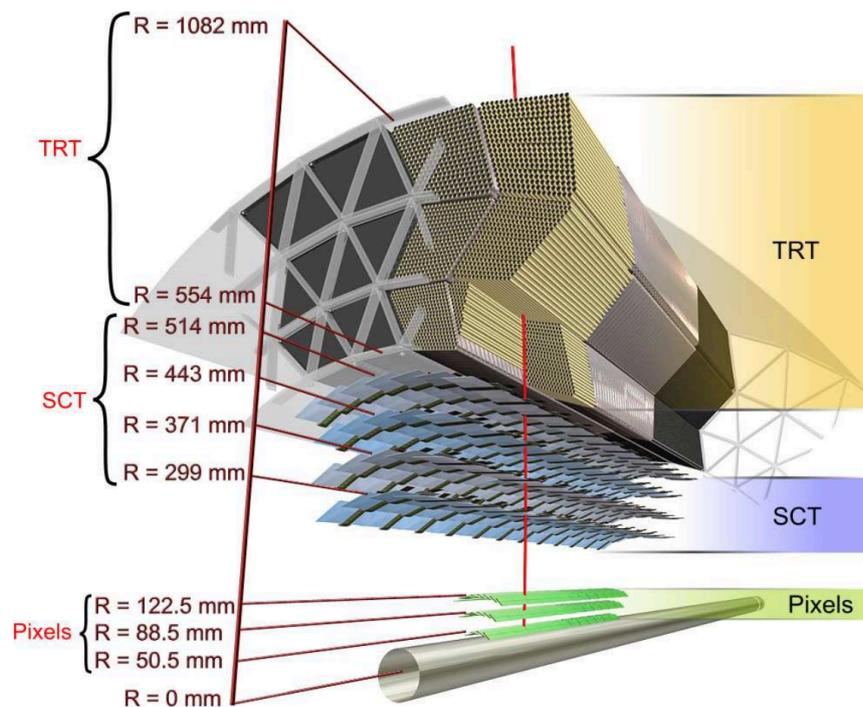


図 2.3: 内部飛跡検出器の構造 [10]

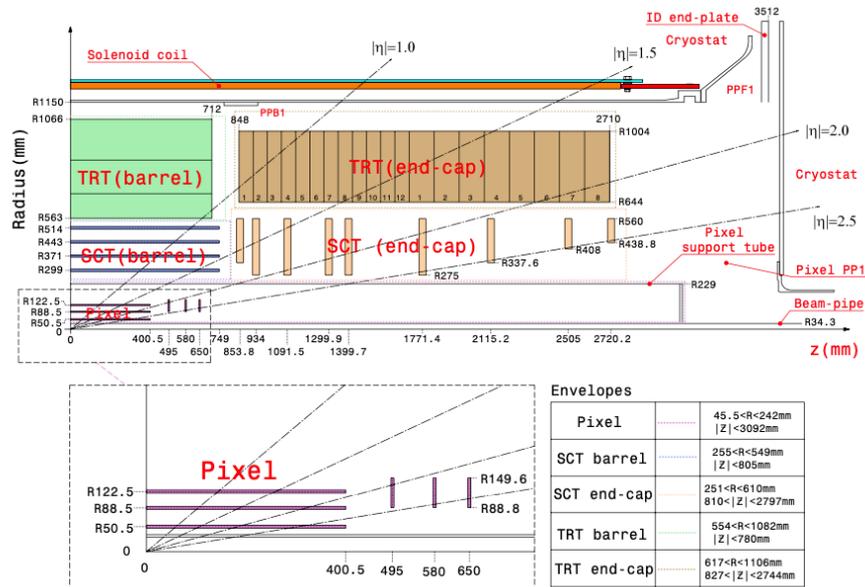


図 2.4: 内部飛跡検出器の断面図 [10]

2.2.3 カロリメータ

カロリメータは内部飛跡検出器の外側に位置する。カロリメータの構造を図 2.5 に示す。内側から電磁カロリメータとハドロンカロリメータの 2 つで構成され、通過した粒子のエネルギーの測定が行われる。エネルギーを落としやすくするために、吸収層には物質量の大きい物質が使用されている。

- 電磁カロリメータ

電磁カロリメータはバレル領域に 1 つ、エンドキャップ領域に 2 つ設置されており、 $|\eta| < 3.2$ の領域を覆っている。鉛の吸収体と液体アルゴンのシンチレータで構成されており、電子と光子による電磁シャワーを利用しエネルギーを測定する。

- ハドロンカロリメータ

原子核の強い相互作用で生じるハドロンシャワーを利用してエネルギーを測定しており、 $|\eta| < 4.9$ の領域を覆っている。バレル領域では鉄の吸収体とプラスチックシンチレータで構成されている。エンドキャップ領域では銅の吸収体と液体アルゴンのシンチレータで構成されている。

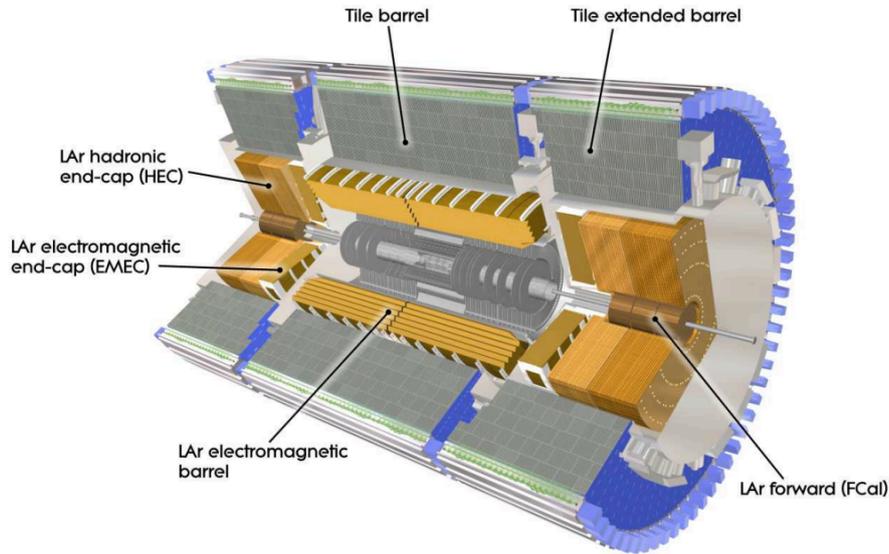


図 2.5: カロリメータの構造 [10]

2.2.4 ミューオン検出器

ミューオン検出器は物質との相互作用が小さいミューオンを検出するためのもので、 $|\eta| < 2.7$ の領域を覆っている。ミューオン検出器は図 2.6 に示すように、荷電粒子の検出をする 4 種類のガス検出器 (RPC, TGC, MDT, CSC) とトロイド磁石で構成されている。トリガーとして用いられているのが Relative Plate Chamber (RPC) と Thin Gap Chamber (TGC) という検出器であり、飛跡の精密測定に用いられているのが Monitored Drift Tube (MDT) と Cathode Strip Chamber (CSC) である。バレル領域に RPC、MDT がエンドキャップ領域に TGC、MDT、CSC が配置されている。

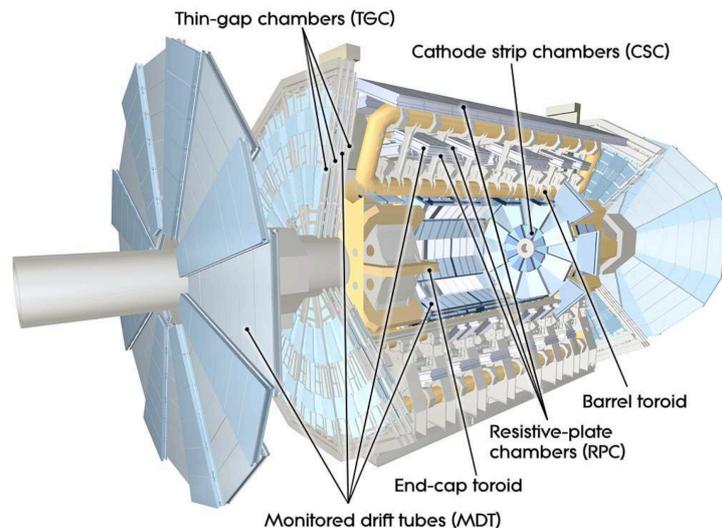


図 2.6: ミューオン検出器の構造 [10]

2.3 トリガーシステム

LHC では 40MHz の頻度で衝突が起こるが、技術的な問題で全ての事象を記録することは不可能である。ATLAS 実験では興味のある必要な事象のみを選別し不要な事象を排除することでレートを抑えており、物理事象に応じて様々なトリガーを用意している。トリガーは 2 段階で行われており、ハードウェアによる高速な判定を行う L1 Trigger(Level 1 Trigger) とソフトウェアによるオンライン事象再構成に基づいた判定を行う HLT(High Level trigger) に分かれる。

- L1 Trigger

L1 Trigger では、40MHz のイベントレートを 100kHz まで落とす。ミュオン検出器とカロリメータの情報のみを用いる。この情報を用いて、高い横運動量をもつミュオン、電子、タウ粒子、光子、ジェット、MET を探す。中央トリガー処理器にて処理され、L1 Trigger を通過したイベントは各検出器のエレクトロニクスと DAQ システムに渡される。

また各イベントにおいて、L1 Trigger は RoI(Region of Interest) という η - ϕ 座標で記述されるコーン状の領域を定義する。RoI データはその領域がもつ特徴の種類と閾値等の基準に関する情報を含み、後段の HLT で使用される。

- HLT

HLT では約 100kHz のイベントレートのデータをオンラインで処理する。L1 Trigger にて定義された RoI において選別が行われる。初めに専用の高速トリガーアルゴリズムを用いてイベントレートを減らし、その後オフラインに近い再構成を行うことで最終的な選別を行っている。全ての検出器情報を用いて RoI 内の飛跡の再構成を行い、事象の選別を行う。選別されたイベントは平均 1.2kHz であり、ストレージに保存される。

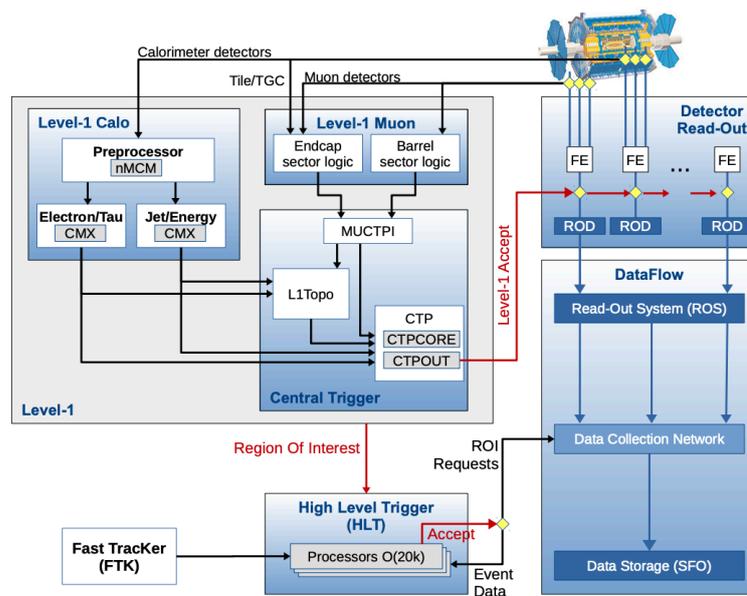


図 2.7: ATLAS トリガーシステムの概略図 [10]

3 タウ粒子

本章では本研究で取り扱うタウ粒子の性質、再構成から同定までの流れを説明する。

3.1 性質

タウ粒子は第三世代の荷電レプトンであり、 -1 の電荷を持つ。反粒子は電荷が $+1$ でその他は同じ性質を持つ。質量は $1776.99 \pm 0.12 \text{ MeV}$ 、平均寿命 τ は $2.90 \times 10^{-13} \text{ sec}$ 、崩壊長 $c\tau = 87.03 \pm 0.15 \mu\text{m}$ である。

タウ粒子の崩壊は大きく分けてレプトン崩壊とハドロン崩壊がある。レプトン崩壊は電子またはミューオンに二つのニュートリノを伴って崩壊する。ハドロン崩壊はハドロンとニュートリノに崩壊する。またタウ粒子の電荷は $+1$ または -1 で電荷は崩壊前後で保存することから、ハドロン崩壊するタウ粒子からは奇数個の荷電粒子が生じる。その生じた荷電粒子の数によって 1prong, 3prong,... と呼ばれる。表 3.1 にタウ粒子の崩壊分岐比を示す。ただし K^\pm も π^\pm に含めて表記している。

表 3.1: タウ粒子の崩壊分岐比

崩壊の種類	終状態	分岐比 [%]
レプトン崩壊	$e\nu\nu$	17.8
	$\mu\nu\nu$	17.4
ハドロン崩壊	$\pi^\pm\pi^0\nu$	25.5
	$\pi^\pm\nu$	10.8
	$\pi^\pm 2\pi^0\nu$	9.3
	$3\pi^\pm\nu$	9.0
	$3\pi^\pm\pi^0\nu$	2.7
	other	7.5

3.2 再構成

ATLAS 実験におけるタウ粒子の再構成の方法について述べる。タウ粒子の崩壊長は $c\tau = 87.03 \pm 0.15 \mu\text{m}$ 、内部飛跡検出器の pixel 検出器の一層目は 33.3 mm のため、ほとんどのタウ粒子は pixel 検出器の一層目に到達する前に崩壊する。よってレプトン崩壊するタウ粒子は電子やミューオンと区別できず、タウ粒子と同定することはできない。またハドロン崩壊するタウ粒子はクォーク由来のジェットとの区別が難しいが、前者は後者よりも細いなどの特徴を用いた機械学習による同定を行っている。ハドロン崩壊するタウ粒子はニュートリノを終状態に含むが、そのニュートリノを検出することはできない。今後タウ粒子についてはニュートリノ以外の検出可能な部分を用いて再構成したものを指す。

シードジェット再構成

タウ粒子はジェットとして再構成されるので、ジェットの再構成法について説明する。ジェットは Anti- k_t アルゴリズム [11] を用いて $\Delta R < 0.4$ で再構成される。ローカルハドロンキャリブレーション (LC) [12] と呼ばれる手法で較正したトポクラスターと呼ばれるカロリメータのセルの 3次元のクラスターを入力として用いる。再構成されたジェットのうち $p_T > 5 \text{ GeV}$ と $|\eta| < 2.5$ を満たすものをタウ候補とし、シードジェットと呼ぶ。

生成点再構成

陽子と陽子の衝突中には複数の相互作用が発生し、複数の生成点が再構成される。多くの粒子はトラックの横運動量の2乗和が最大になるような点を生成点とし、その生成点を primary vertex と呼ぶ。しかしタウ粒子は崩壊前に primary vertex から離れるため、正しい生成点を特定するために TJVA[13] と呼ばれるアルゴリズムが用いられている。最も f_{p_T} が大きくなるような点をタウの生成点とする。 f_{p_T} はシードジェットの $\Delta R < 0.2$ のすべてのトラックを用いて式 (3.1) で計算される。

$$f_{p_T} = \frac{\Sigma p_T(\text{生成点に関連づけられたトラック})}{\Sigma p_T(\text{全てのトラック})} \quad (3.1)$$

正しい生成点を選択する効率について 1prong と 3prong のそれぞれで平均 pp 衝突回数 $\langle \mu \rangle$ とタウ粒子の横運動量 p_T の関数として図 3.1 に示す。この生成点再構成アルゴリズムを用いると良い効率で正しい生成点を再構成でき、とくに低横運動量領域で顕著である。

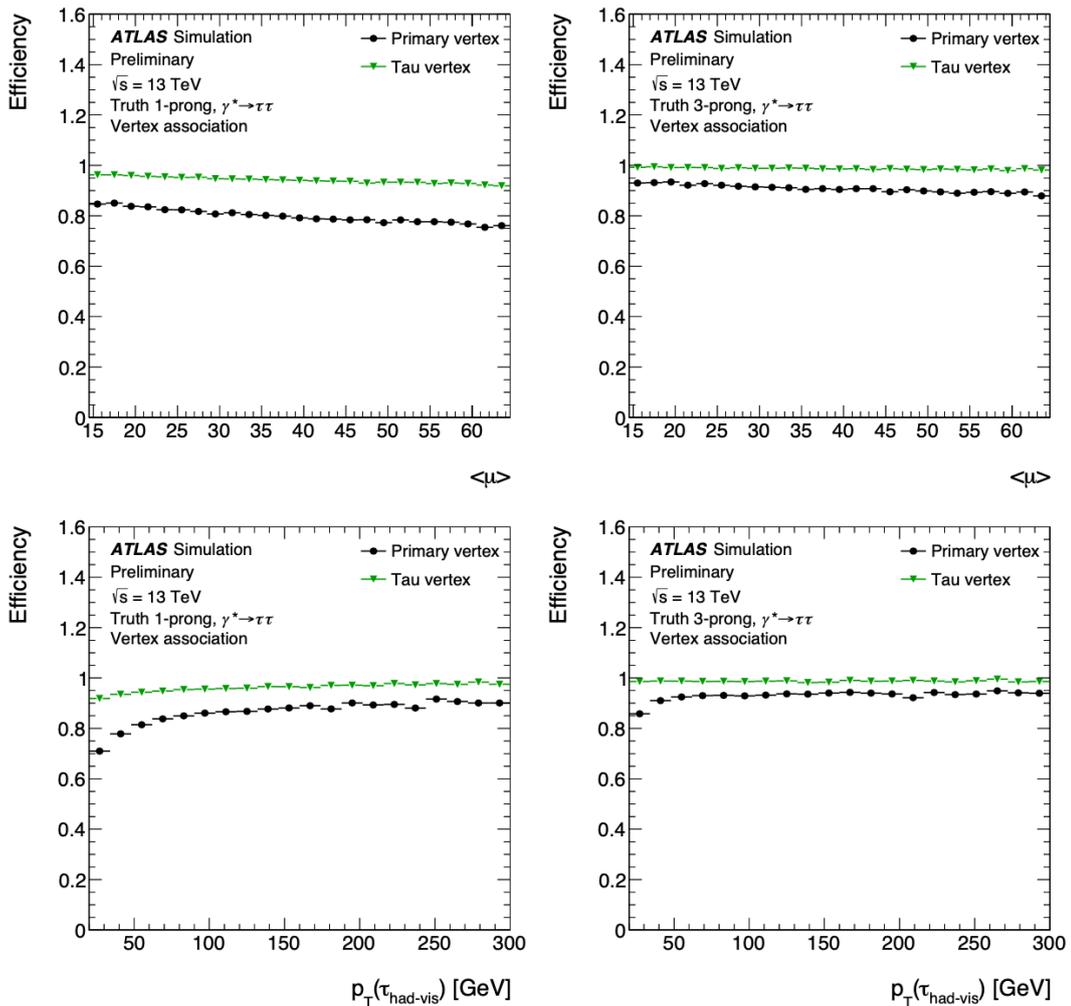


図 3.1: 生成点再構成効率 [15]

3.3 トラック識別

トラックはタウ粒子の $\Delta R < 0.25$ にある場合、および次の基準を満たす場合に関連づけられる。

- $p_T > 1\text{GeV}$
- pixel 検出器に少なくとも二つのヒット
- pixel 検出器と SCT 検出器合わせて少なくとも七つのヒット
- $|d_0^{TJV A}| < 1.0\text{mm}$
- $|z_0^{TJV A} \sin \theta| < 1.5\text{mm}$

$0.25 < \Delta R < 0.4$ のトラックに関しては、シードジェットに対して ghost-particle association technique[14] という手法によって一致するか、 ΔR が一番小さいジェットがタウ粒子のジェットである場合、関連づけられる。

タウ粒子の電荷と荷電粒子の数を正確に決定するために、タウ粒子の崩壊によって生じたトラックとそれ以外を区別する必要がある。次の 4 つのカテゴリにトラックを分類する。

- タウトラック
タウ粒子の崩壊によって生じた荷電粒子。
- 対生成トラック
タウ粒子の崩壊粒子の $\pi^0 \rightarrow \gamma(\rightarrow ee)\gamma(\rightarrow ee)$ によって生じる電子。
- 孤立トラック
クォークやグルーオン由来のジェットから生じるトラック。これらのトラックは低い横運動量で、広く生じる傾向にある。
- フェイクトラック
誤ったヒットを結んでできたトラックまたはパイルアップ由来のトラック。

分類は Recurrent Neural Network(RNN) というアルゴリズムを用いた機械学習で行っている。信号事象として $\gamma^* \rightarrow \tau\tau$ 、背景事象として di-jet の MC サンプルの $p_T > 5.5\text{GeV}$ のタウ粒子を用いる。表 3.2 に機械学習に用いている入力変数の一覧を示す。また図 3.2 に truth 1prong と truth 3prong それぞれのトラック分類の効率について示す。横軸は正しいトラックのタイプを表していて、縦軸は RNN によって分類されたトラックタイプを示している。オレンジの部分が正しく分類できた効率で、truth 1prong のタウトラック (TT) は 95.2 % の効率でタウトラックと認識されることを示す。また図 3.3 に truth 1prong と truth 3prong のタウ粒子がいくつの prong 数に分類されたかの効率を示している。truth 1prong は約 90 % の確率で 1prong と分類され、truth 3prong は約 70 % の確率で 3prong と分類される。

表 3.2: トラック識別で使用している入力変数

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

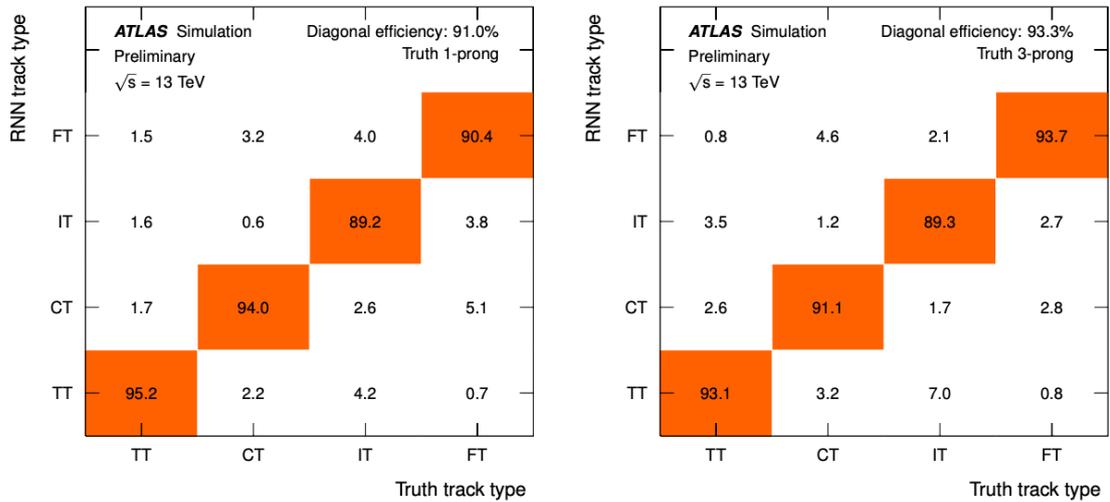


図 3.2: 信号事象のトラック識別効率 (左は Truth 1-prong, 右は Truth 3-prong)[15]

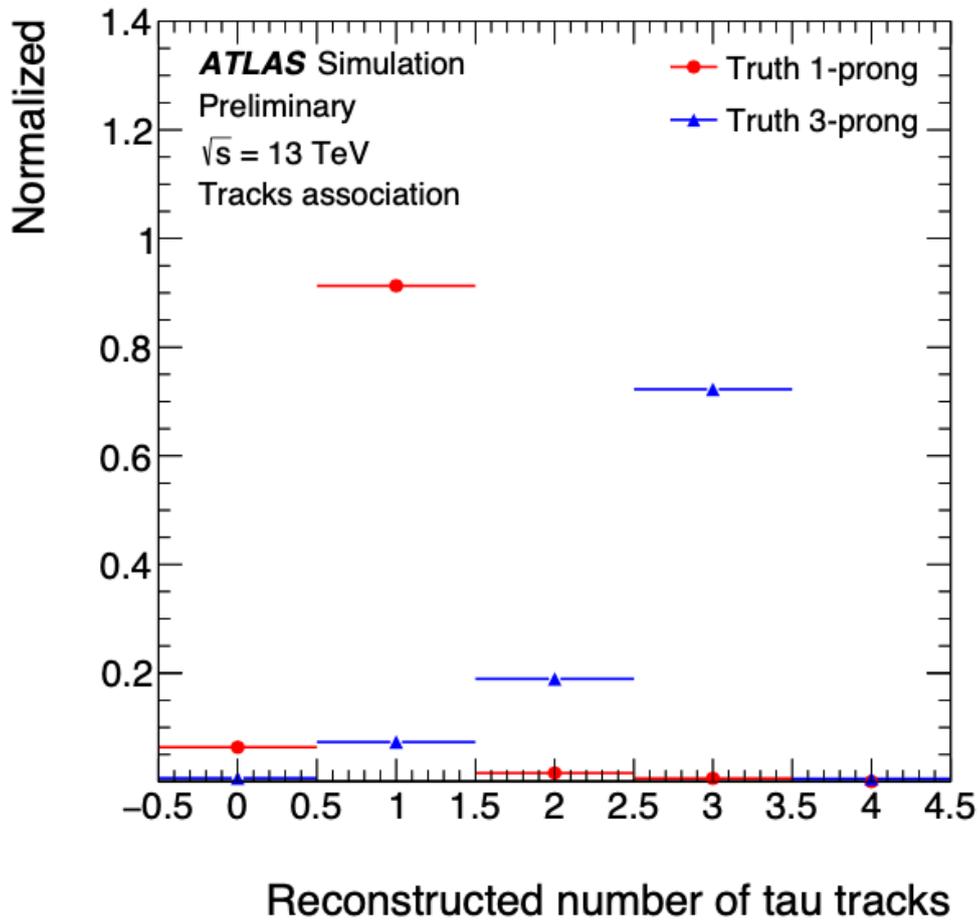


図 3.3: Truth 1-prong, Truth 3-prong それぞれにおけるタウトラックの本数の割合 [15]

3.4 同定

ジェットとして再構成されたタウ粒子候補の中には、タウ粒子だけではなくクォークやグルーオン由来の QCD ジェットも含まれている。新粒子探索の感度向上のためにも QCD ジェットの背景事象を排除し効率よくタウ粒子を取得する必要がある。そのために Recurrent Neural Network(RNN) というアルゴリズムを用いた機械学習でタウ粒子の同定が行われている。信号事象として $\gamma^* \rightarrow \tau\tau$ 、背景事象として di-jet の MC サンプルの $p_T > 20\text{GeV}$ のタウ粒子を用いる。1prong、3prong の特徴の違いから別々に学習する。表 3.3 に機械学習に用いる入力変数の一覧を示す。入力変数はジェット、ジェットがカロリメータに残したクラスター、ジェットを構成するトラックの 3 つの情報を用いる。

図 3.4 にタウ粒子同定の結果を示す。横軸は信号事象の取得効率を示していて、縦軸は背景事象を何分の一に減らせるかを示している。図中にある点は表 3.4 に示す信号事象取得効率によって定められている基準点である。

表 3.3: タウ粒子同定で使用している入力変数

変数	説明	1prong	3prong
$p_T^{\tau_{had-vis}}$	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	✓	✓
$z_0^{TJVA} \sin\theta$	Tau vertex longitudinal impact parameter	✓	✓
d_0^{TJVA}	Tau vertex transverse impact parameter	✓	✓
$S(z_0^{TJVA} \sin\theta)$	Tau vertex longitudinal impact parameter divided by its uncertainty	✓	✓
$S(d_0^{TJVA})$	Tau vertex transverse impact parameter divided by its uncertainty	✓	✓
$p_T^{\tau_{had-vis}}$	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	✓	✓
$\langle \lambda_{centre} \rangle$	Mean distance of the cluster shower centre from the calorimeter front face measured along the shower axis	✓	✓
$\langle \lambda^2 \rangle$	Mean distance of a cell from the shower centre along the shower axis	✓	✓
$\langle r^2 \rangle$	Second moment in the radial distance of cluster cells from the cluster axis	✓	✓
$p_T^{\tau_{had-vis}}$	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		✓

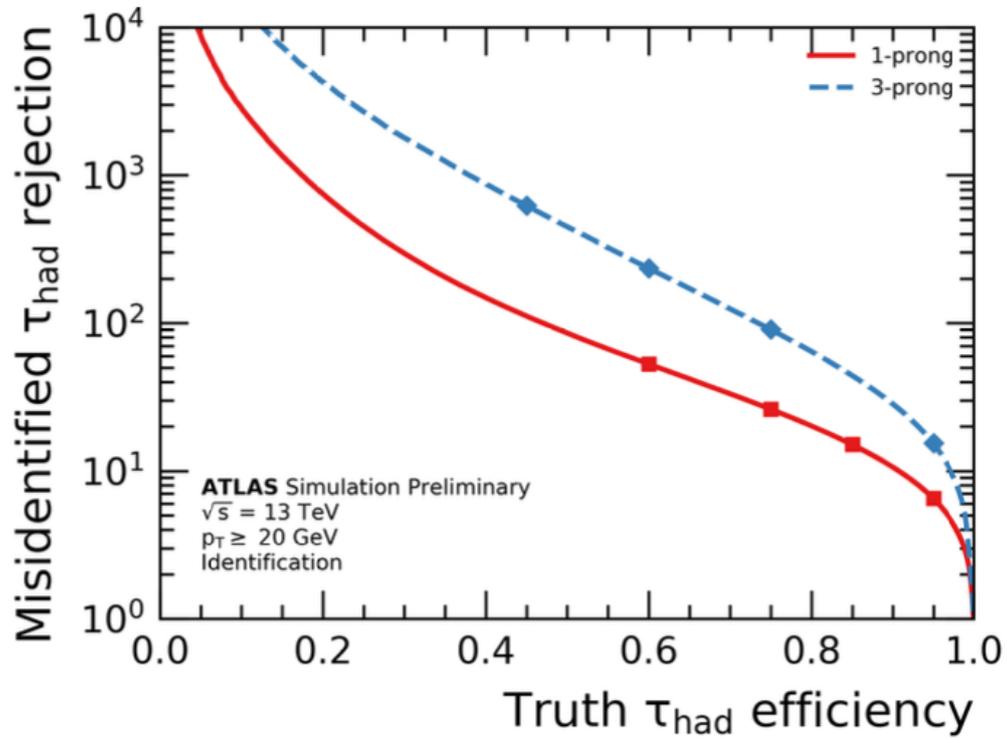


図 3.4: タウ粒子同定性能
(背景事象取得効率の逆数を信号事象の取得効率の関数であらわしたもの)[15]

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 %

4 グラフニューラルネットワークを用いたタウ粒子同定

タウ粒子同定はカットベースな手法から BDT(決定木)、RNN(Recurrent Neural Network) と、より発展した手法を用いてきた。本研究では新たにグラフニューラルネットワークを用いた手法を開発した。まずはグラフニューラルネットワークについて説明する。

4.1 グラフニューラルネットワーク

グラフニューラルネットワーク (GNN) はグラフデータを用いることを得意とした深層学習である。グラフデータ (G) は頂点 (V) と辺 (E) で表されるものであり、 $G=\{V,E\}$ と表されることが多い。頂点のことをノード、辺のことをエッジとも呼ぶ。図に示す具体例は、ノードの集合が $V=\{0,1,2,3\}$ 、エッジの集合が $E=\{(0,1),(0,2),(2,1),(2,3)\}$ となるグラフである。

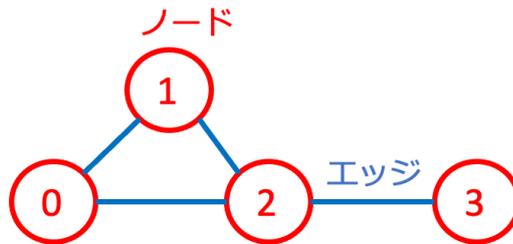


図 4.1: グラフの例

GNN で解決できる問題

GNN で解決できる問題としてノード分類、リンク予測、グラフ分類の 3 つに分けられる。図 4.2 にそれぞれのイメージ図を示す。

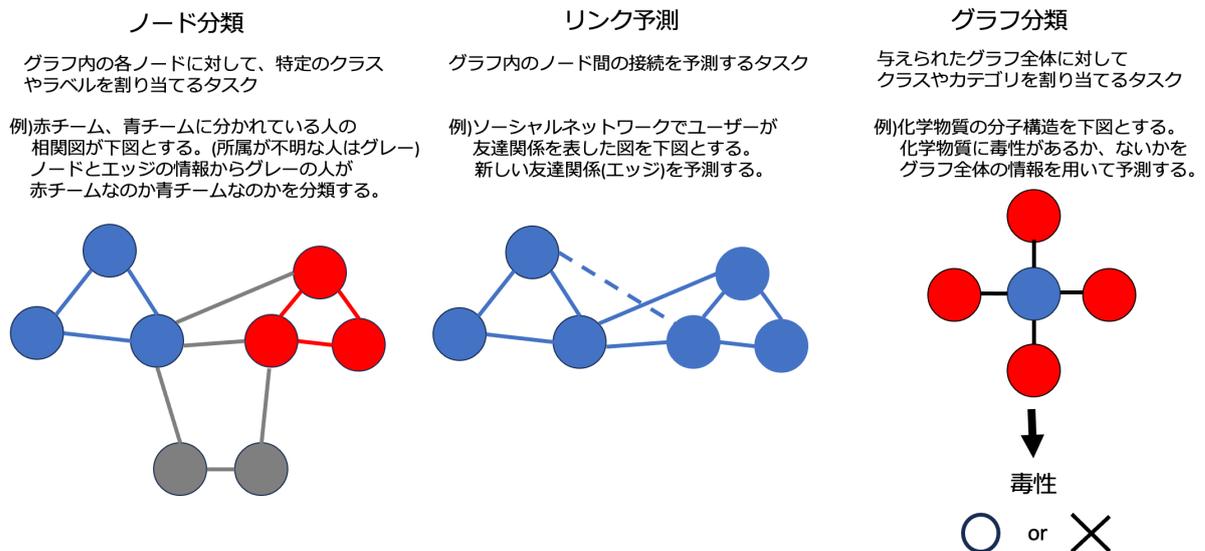


図 4.2: GNN で解決できる問題

ジェットに付随するトラック、クラスターを頂点とするグラフを用いたグラフ分類によって、タウ粒子か QCD ジェットのどちらに属するかを予測する。

GNN の種類

多くの GNN では入力データとなるグラフを GNN の層に通し、出力として得られるノード、エッジ、グラフの埋め込み表現をその後の学習や予測に用いる。図 4.2 に GNN の一般的な構成についての図を示す。図 4.2 の GNN layer に相当する部分は様々なものが提案されており、分類したものを図 4.3 に示す。特に今回使用している GraphSAGE という手法について説明する。

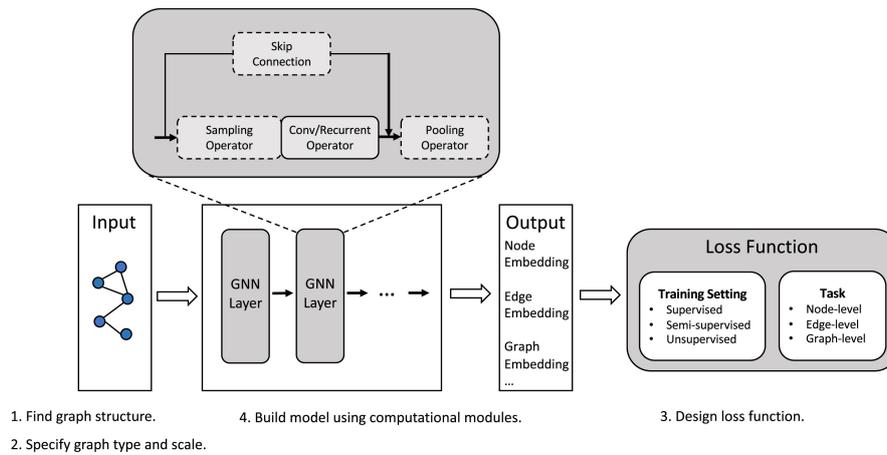


図 4.3: GNN の学習の流れ [16]

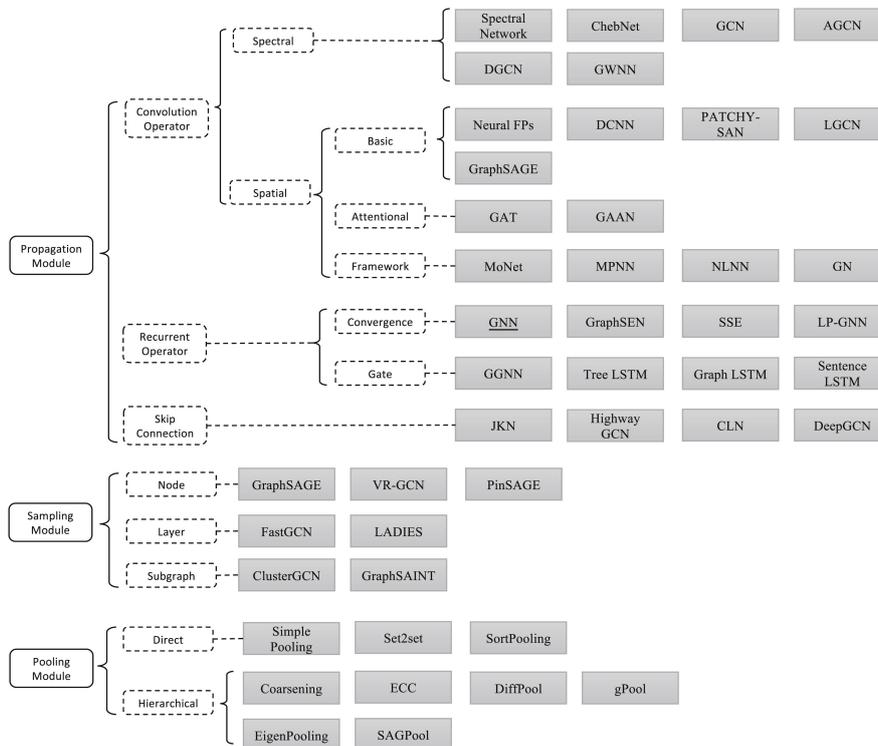


図 4.4: GNN のモデルの種類 [16]

GraphSAGE

GraphSAGE は図 4.3 にあるように Spatial な手法である。Spatial な手法とはある頂点を中心に見た時に、その近傍の頂点の情報を利用するものであるが、いくつか問題点がある。その一つが近傍爆発というもので、畳み込みの層を深くしていくと一つの頂点のために情報を集約する必要のある近傍の数は深さに応じて指数関数的に増えていくという問題である。また使う頂点数が多くなりすぎると、集約結果がどの頂点でも似たようなものになってしまう過剰平滑化という問題もある。

上記の問題を解決する手法の一つとして GraphSAGE がある。図 4.4 に演算の流れを示す。各ノードについて個別の埋め込みを学習するかわりに、イテレーションごとに畳み込みに利用する頂点をサンプリングし集約関数を学習する。サンプリングにより、次数が大きい頂点に関して計算量を一定に抑えることができるというメリットがある。

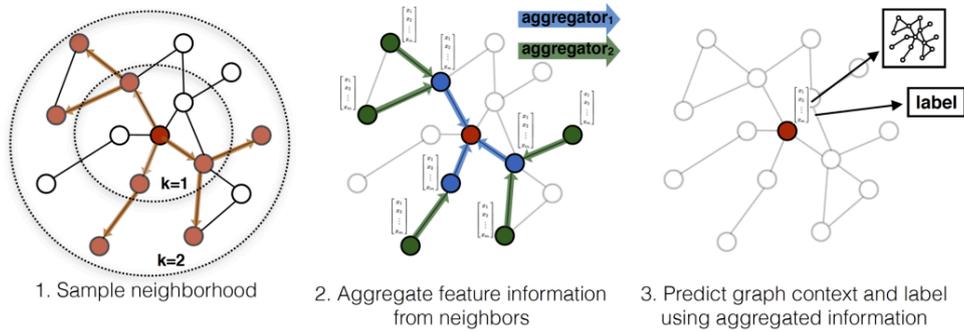


図 4.5: GraphSAGE の演算の流れ [17]

4.2 学習・評価方法

GNN を用いたタウ粒子同定の学習・評価方法について説明する。

使用したサンプル・カット

信号事象として $\gamma^* \rightarrow \tau\tau$ 、背景事象として di-jet の MC サンプルを用いる。信号事象、背景事象ともにタウ粒子として再構成されたオブジェクトの数はイベントごとに異なり、複数存在する場合もある。したがってデータサンプル数における単位はイベント数ではなく、タウ粒子として再構成されたオブジェクトの数とする。表 4.1 にタウ粒子として再構成されたオブジェクトのカット条件を示す。 p_T と $|\eta|$ のカットは信号事象、背景事象共に同じものを使用している。1prong、3prong の特徴の違いから別々に学習している。信号事象は MC の truth 情報と再構成された prong 数が一致しているという条件を追加している。

表 4.1: サンプルのカット条件

	信号事象	背景事象
p_T	$> 20 \text{ GeV}$	
$ \eta $	< 2.5	
	1prong and truth 1prong	1prong
N_{track}	or	or
	3prong and truth 3prong	3prong

入力変数

入力変数は表 3.3 に示している現行手法と同じものを用いる。信号事象と背景事象の違いを説明するために、いくつかの変数の分布を示す。図 4.5 に dR_{max} と $centFrac$ という変数の分布を示す。 dR_{max} はタウジェットとタウジェットに付随するトラックの ΔR の最大値であり、信号事象の方が細く分布していることが分かる。これはタウ粒子はほとんどが pixel 検出器に到達する前に崩壊し、質量が約 1.77GeV と軽いため、崩壊粒子の運動量も大きくなるからである。 $centFrac$ はジェットの中心に落としたエネルギーの割合であるが、信号事象の方がより中心にエネルギーを落としていることが分かる。これは dR_{max} 分布を見ると分かるように、信号事象の方が細い傾向にあることに由来する。これらは一例であるが、他の入力変数を含めた信号事象と背景事象の違いを多変量解析し、区別するために機械学習が用いられている。

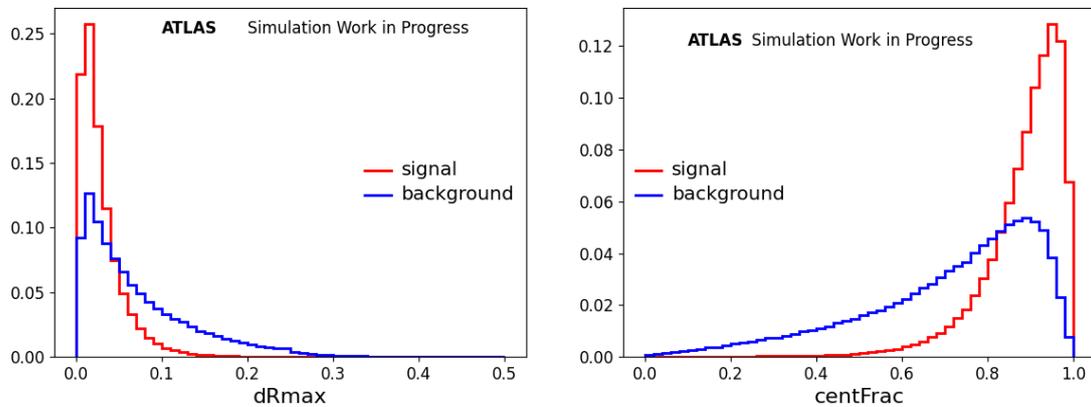


図 4.6: 信号事象と背景事象の入力変数分布の違いの例

モデル

学習に用いたモデルについて説明する。図 4.7 に採用した GNN の学習モデルを示す。これは現行手法のモデルを参考にしている。トラックとクラスターとジェットの情報を別々に学習させた後、情報を結合して学習を進め、最後は 0 以上 1 以下のスコアを算出している。このスコアはこれから GNNscore と呼ぶことにする。GNNscore は 0 に近いほど背景事象らしく、1 に近いほど信号事象らしいスコアである。

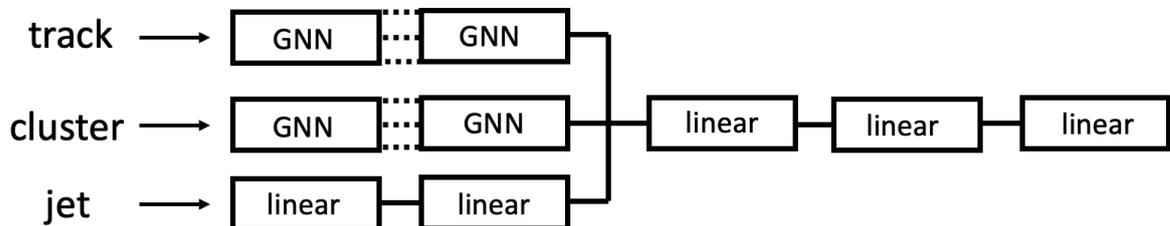


図 4.7: 採用した GNN の学習モデル

トラックとクラスターはタウ粒子として再構成されたオブジェクトに対して複数存在するので、GNN を用いて学習をする。図 4.7 の GNN は前述の GraphSage という手法のことであり、何層にするのかという調整については後ほど説明する。図 4.8 にトラックの入力のイメージ図を示す。再構成されたトラックが 4 本あったとき、それをノードとして全結合でグラフを作成する。クラスターも同様にグラフを作成する。トラックは p_T の降順で 10 本、クラスターは E_T の降順で 6 つを上限としている。

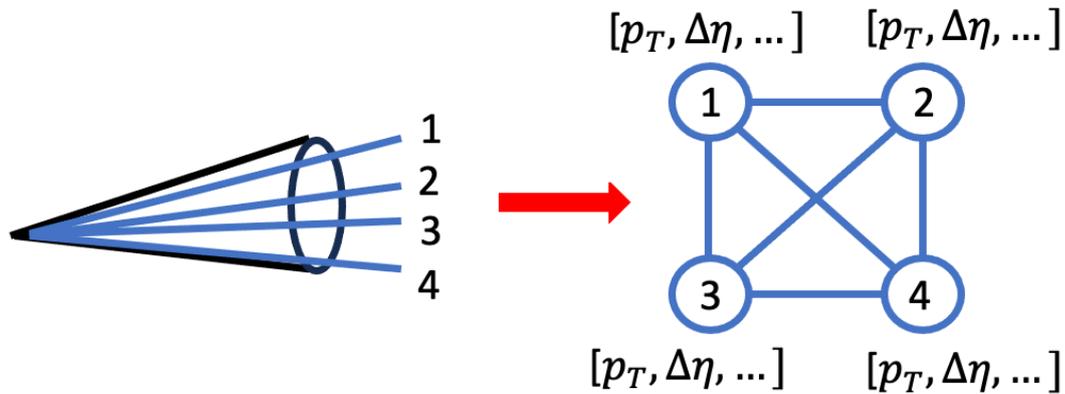


図 4.8: 入力変数のグラフ化のイメージ図

図 4.7 の linear は線形層のことである。線形層は入力ベクトルを X 、重み行列を W 、 b をバイアスベクトルとした場合、出力 Z は $X \cdot W + b$ で表される。図 4.9 に線形層のイメージ図を示しており、3 次元の入力が重み W によって 4 次元に変換されている。

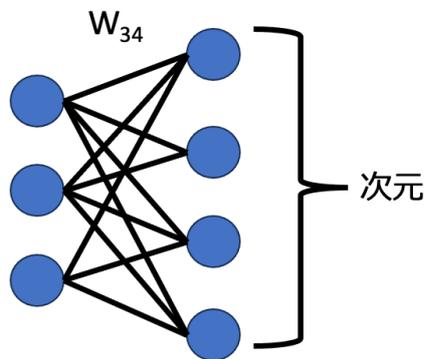


図 4.9: 線形層のイメージ図

学習・評価

ここまでサンプル・カット、入力変数、学習モデルについて説明してきた。それらを用いた学習・評価の流れについて説明する。まず学習用のトレーニングデータ、過学習を起こしていないか確認する検証用のバリデーションデータ、評価用のテストデータの 3 つに分ける。それぞれのデータは信号事象と背景事象のデータサンプル数が 1:1 で構成されるようにする。過学習とはトレーニングデータに対して過剰に適応され、未知のデータに対しては性能が低くなる現象である。

学習時にはトレーニングデータとバリデーションデータを用いて、以下の流れで行う。

1. データの前処理を行う。
2. トレーニングデータをバッチに分ける。
3. 1 バッチずつ学習させる。
4. バリデーションデータで検証する。
5. 3 と 4 を繰り返す。

1については、入力変数によって値の大きさが異なるため、標準偏差と平均値を用いて0から1に規格化している。2については、一度にすべてのデータを読み込まず、バッチにして細かく読み込むことでメモリ消費を抑えるなどの利点がある。3については、バッチごとに損失というものを計算し、その損失を逆伝播させることで重みを更新し、学習を進めている。損失の計算には Binary Cross Entropy という関数を用いていて、式 (4.1) で表される。

$$\text{loss}(y, p) = -[y \log(p) + (1 - y) \log(1 - p)] \quad (4.1)$$

y は真のラベルであり、信号事象の場合 1、背景事象の場合 0 である。図 4.7 の学習モデルの最終層で出力される GNNscore が信号事象である確率 p であり、背景事象である確率は $(1-p)$ である。この真のラベルとの誤差である損失を減少させていくことが重要である。4については、3と同じようにバリデーションデータで損失を計算する。しかし検証用のデータであるので、損失は逆伝播させない。3と4の過程を一回行うことをエポックと呼び、損失が減少しなくなるまで同じデータで学習と検証を繰り返す。図 4.10 に学習中の損失の推移の例を示す。エポックが増えるにつれて学習が進み、損失が減少していることが分かる。図中の実線はトレーニングデータの損失の推移、点線はバリデーションデータの損失の推移である。前者の損失のみ減少し、後者の損失は減少しないという過学習が起きていないか確認しながら学習を進めている。今回は 10 エポック連続でバリデーションデータの損失を更新できなければ学習を終了している。

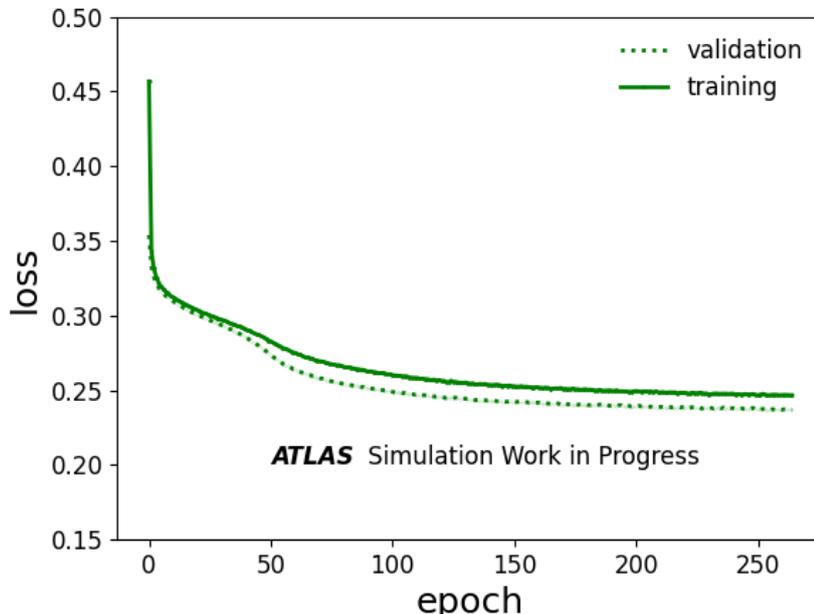
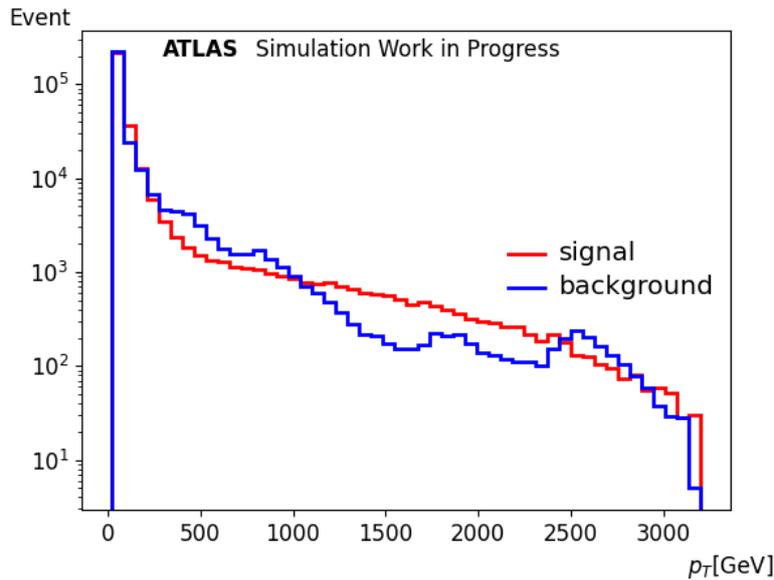


図 4.10: 学習中の損失の推移

また学習時に行っていることがもう 1 点ある。それは横運動量 p_T による重み付けである。図 4.11 に 1prong の信号事象と背景事象の p_T 分布を示す。信号事象、背景事象ともに低い横運動量領域のデータサンプル数が多く、高い横運動量領域のデータサンプル数が少ない。このまま学習を行うと低い横運動量領域の特徴を学習し過ぎてしまい、高い横運動量領域のパフォーマンスが落ちてしまうことが懸念される。重みは式 (4.1) の損失に掛け算し、低横運動量領域には 1 より小さい重み、高横運動量領域には 1 より大きい重みを与えることで上記の問題を緩和している。

図 4.11: 横運動量 p_T 分布

次に評価方法について説明する。テストデータは信号事象は truth 1-prong かつ 1prong、背景事象は 1prong のものを用いる。3prong も同様である。まず学習済みモデルを用いて、テストデータの GNNscore を算出する。図 4.12 に GNNscore 分布の例を示す。信号事象 (赤) は 1、背景事象 (青) は 0 にピークをもって分布しており、信号事象と背景事象の分類がされていることが分かる。次に、ある GNNscore を閾値としてその値より大きいイベントを信号事象、小さいイベントを背景事象としたときに、信号事象の取得効率と背景事象除去能力 (背景事象が何分の一になるか) を計算する。GNNscore の閾値を変えながら信号事象の取得効率を横軸に、背景事象除去能力を縦軸にとったグラフを図 4.13 に示す。より高い信号事象取得効率とより高い背景事象除去能力、つまりグラフがより右上に描かれるような学習モデルを作成する。

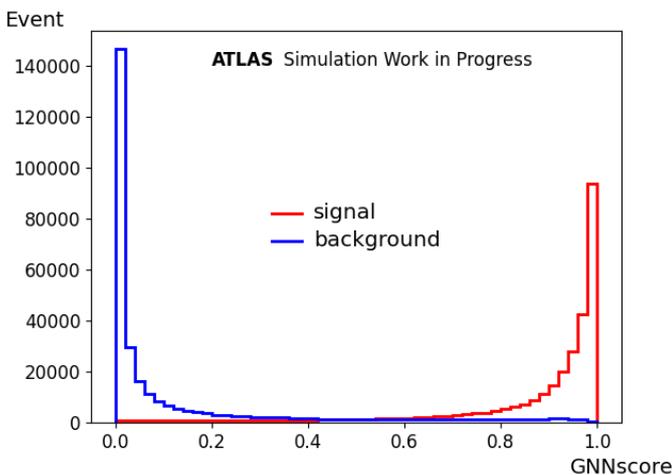


図 4.12: GNNscore 分布の例

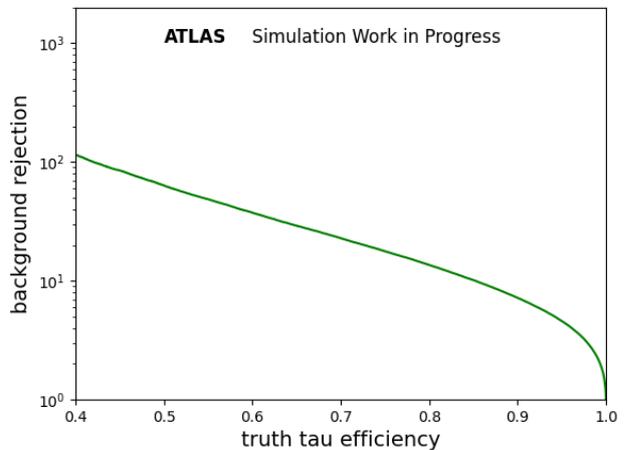


図 4.13: タウ粒子同定の性能評価の例

4.3 パラメータ調整

タウ粒子同定の性能を上げるために行ったパラメータ調整について説明する。図 4.7 の採用した GNN の学習モデルの GNN 層の数について調整をした。また GNN 層の中には図 4.9 に示している線形層が内包されており、その次元も調整した。機械学習では一般的にモデルのパラメータ数が大きいほど性能は向上し、そのパラメータ数はデータサンプル数に制限される。パラメータ数は層数や次元を増やすと大きくなるので、どちらを増やすのが効果的か調べた。データサンプル数を 480k 用いて、(次元、層数)を (64,2),(128,2),(64,3),(128,3) と変化させたときの結果を図 4.14 に示す。次元は 64 より 128、層数は 2 より 3 と大きくした方が性能が上がっていることが分かる。表 4.2 に基準点 Medium (efficiency 75 %) のときの結果を示す。(128,2) と (64,3) を比べたとき、後者の方がパラメータ数が小さいにも関わらず背景事象除去能力は約 1.4 倍になっていることが分かる。したがって次元を大きくするよりも層数を増やした方が性能向上に効果的である。

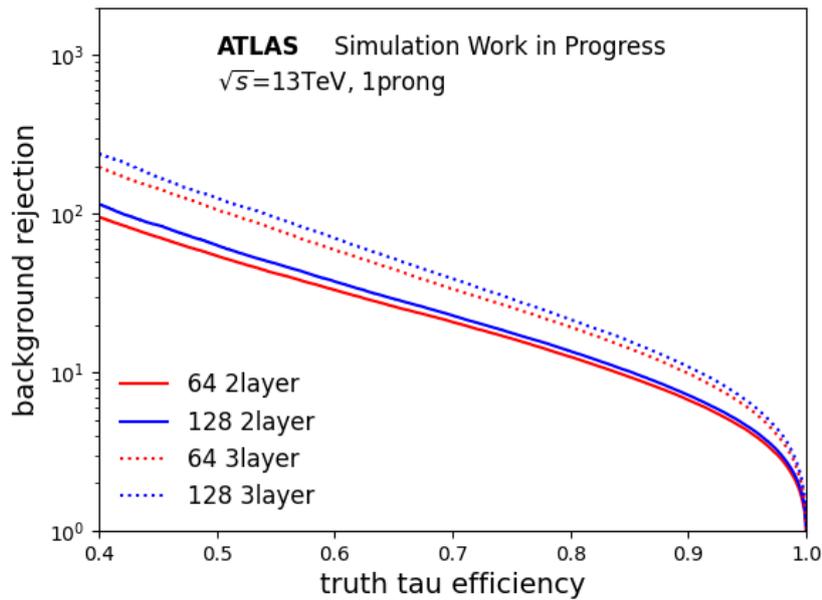


図 4.14: 層数と次元を増やしたときの性能変化 (1prong)

表 4.2: 基準点 Medium における層数と次元を増やしたときの性能変化

(次元、層数)	パラメータ数	背景事象除去能力
(64,2)	42k	16.4
(128,2)	141k	17.8
(64,3)	57k	25.7
(128,3)	207k	29.5

次はデータサンプル数を増やしたときの性能変化について説明する。学習の際は前述のパラメータ調整の結果などを用いて、過学習を起こさず性能が最大化するようにパラメータ数を決めている。データサンプル数を 480k, 1.6M, 3.2M と増やしたときのパラメータ数と基準点 Medium のときの背景事象除去能力の変化について図 4.15 に示す。データサンプル数を増やすとパラメータ数も大きくなり、背景事象除去能力も向上していることが分かる。データサンプル数が 480k のときと 3.2M のときを比べると、30% 以上向上している。

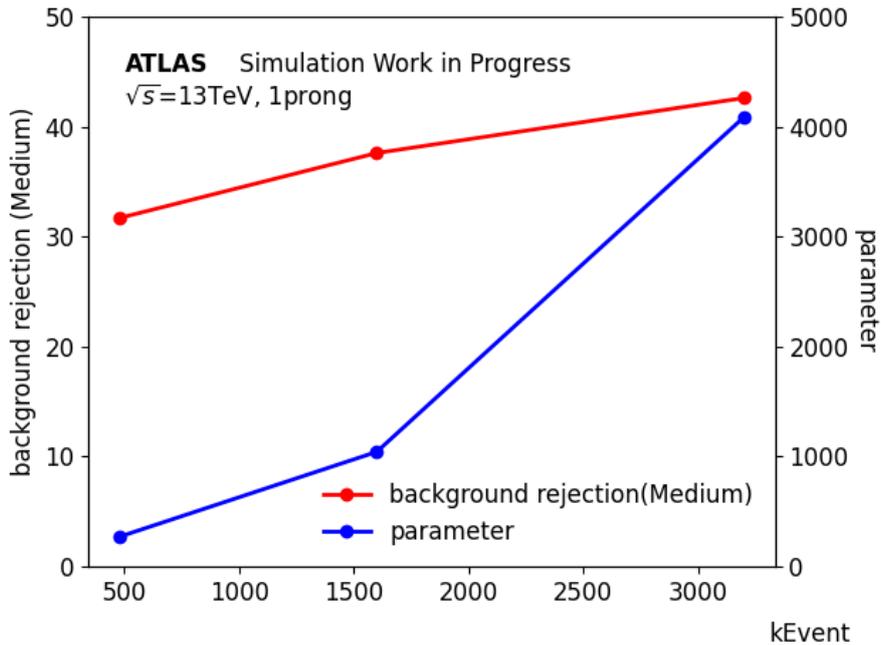


図 4.15: データサンプル数を増やした時のモデルのパラメータ数とそのときの背景事象除去能力 (Medium) の変化

4.4 現行手法との比較

これまで説明してきた GNN を用いた手法と RNN を用いた現行手法の性能を比較する。図 4.16 に 1prong、3prong それぞれにおける結果を示す。GNN を用いた手法では 1prong は 3.2M、3prong は 1.6M のデータサンプルを用いて学習を行った。1prong、3prong ともに現行手法より 1.x 倍のファクターで性能が良いことが分かる。前述の結果よりデータサンプル数を増やすことで、より現行手法より性能が向上することが期待される。

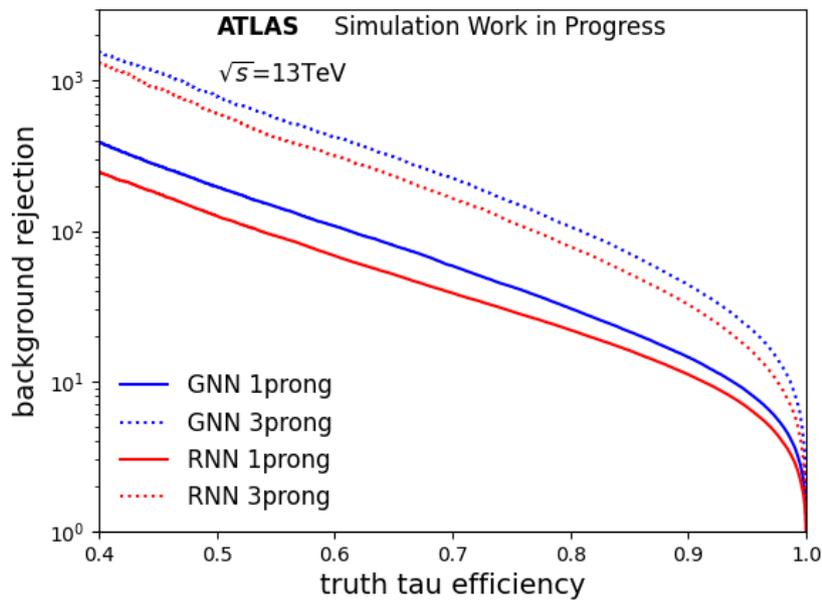


図 4.16: 現行手法と GNN を用いた手法のタウ粒子同定性能比較

横運動量領域での比較

1.2 項タウ粒子を用いた解析で説明したように、横運動量が低いタウ粒子から高いものまで必要とされている。よって low(15GeV~30GeV)、middle(30GeV~100GeV)、high(100GeV~) の 3 つの領域に分けて現行手法と比較する。図 4.17 に low,middle,high の結果を示し、表 4.3 に基準点 Medium における背景事象除去能力を示す。1prong,3prong ともに横運動量領域に関係なく背景事象除去能力が向上していることがわかる。

表 4.3: 基準点 Medium における横運動量領域ごとの背景事象除去能力の比較

	背景事象除去能力 (low)	背景事象除去能力 (middle)	背景事象除去能力 (high)
RNN 1prong	11.9	31.4	106.6
GNN 1prong	17.8	39.4	143.3
RNN 3prong	81.9	304.3	1282
GNN 3prong	112.5	402.7	1754

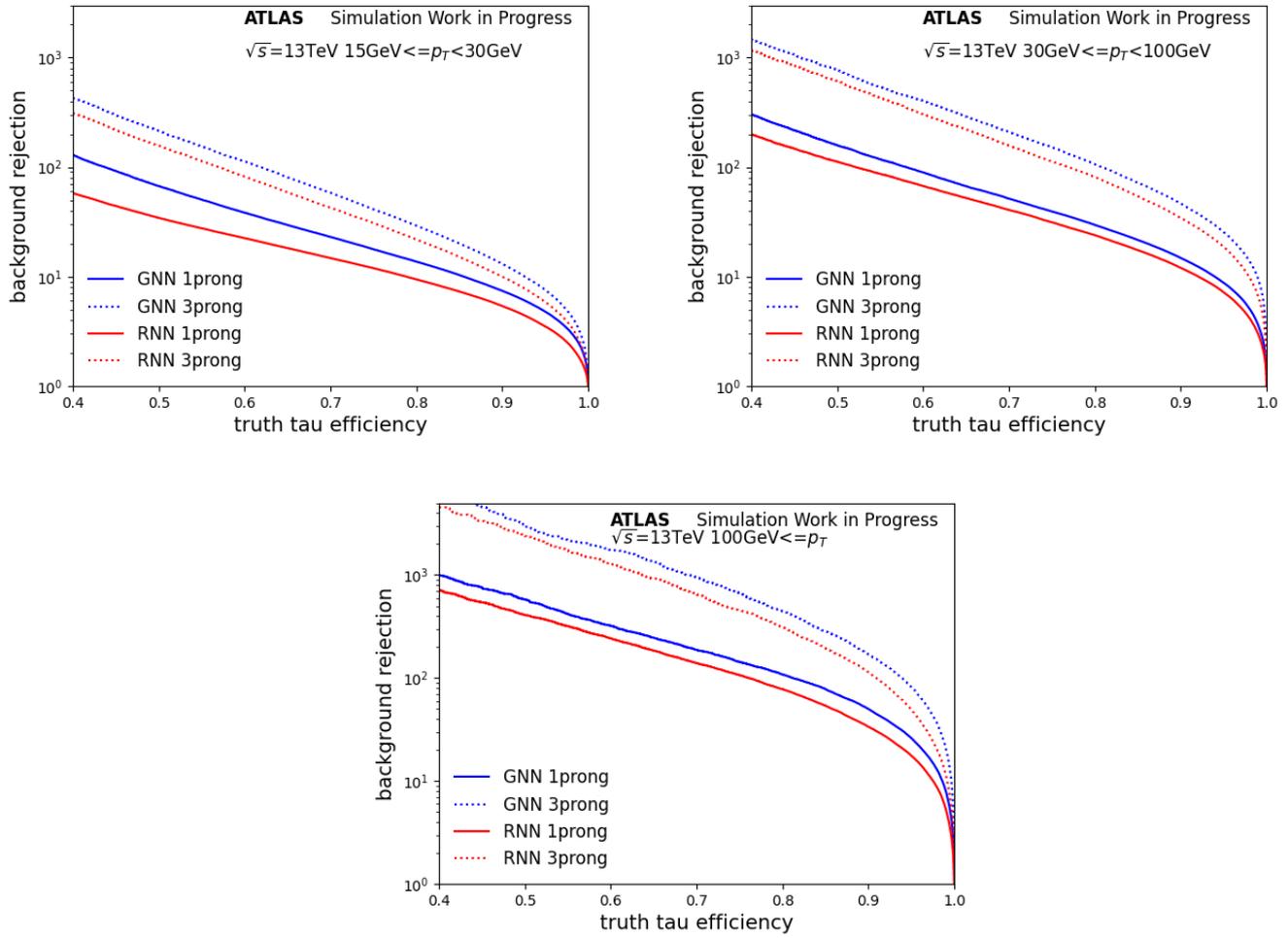


図 4.17: 横運動量領域ごとの背景事象除去能力の比較

5 トラック識別も含めたタウ粒子同定

よりタウ粒子同定の性能を向上させるために新しい学習モデルを検討した。現行手法と4章のGNNを用いた手法と新たに検討した手法の違いについて、図5.1に簡単に示す。3章でも説明したが、タウ粒子同定の前にはトラック識別が行われており、prong数などが決定されている。このトラック識別の性能を向上させ信号取得効率を上げることが重要であり、トラック識別もGNNを用いることで改善を図る。また深層学習の強みは大規模なモデルであり、一度に大量のデータを学習させることが好ましいため、トラック識別とタウ粒子同定を同時に行うことを検討する。トラック識別における学習内容がタウ粒子同定にも活かされることを期待する。

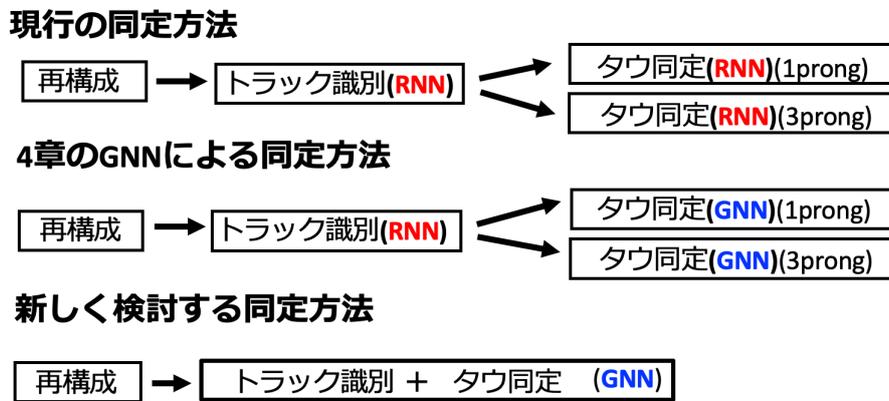


図 5.1: 新たな学習方法の検討

5.1 学習方法

使用したサンプル・カット

使用したサンプルは4章の手法と同じく、信号事象として $\gamma^* \rightarrow \tau\tau$ 、背景事象として di-jet の MC サンプルを用いる。トラック識別も含めるにあたり、カットの条件を変更したのでそれを表5.1に示す。横運動量のカットは現行のトラック識別で用いられている閾値に合わせた。またトラック識別により荷電トラックの数、prong数が決定するので、prong数の指定は行わない。

表 5.1: サンプルのカット条件の変更点

	変更前		変更後	
	信号事象	背景事象	信号事象	背景事象
p_T	> 20 GeV		> 5.5 GeV	
$ \eta $	< 2.5		< 2.5	
N_{track}	1prong and truth 1prong or 3prong and truth 3prong	1prong or 3prong	—	—

入力変数

入力変数は表3.3に示しているタウ粒子同定の変数に追加して、表3.2に示しているトラック識別の変数を用いる。

モデル

学習に用いたモデルについて説明する。図 5.1 にトラック識別も含めた学習モデルを示す。黒で書かれている部分は 4 章で用いていたタウ粒子同定のモデルであり、赤で書かれている部分はトラック識別を行うモデルである。トラック識別はトラックをインプットとした GNN 層の出力を用いて行う。

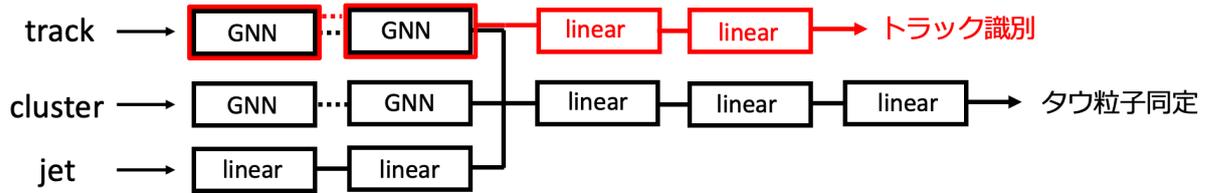


図 5.2: トラック識別も含めたタウ粒子同定の学習モデル

学習

学習時の流れは 4 章で説明したものと同じである。損失の計算方法に変更があるので、それについて説明する。図 5.1 のモデルの図に示すように、タウ粒子同定、トラック識別の 2 つの出力があり、それぞれで損失を計算する。前者は Binary Cross Entropy という関数を用い、後者は Cross Entropy という関数を用いる。Cross Entropy は次の式 (5.1) で表される。

$$loss(y, p) = -\sum_x y(x) \log(p(x)) \quad (5.1)$$

トラック識別はタウトラック、対生成トラック、孤立トラック、フェイクトラックの 4 つを識別するタスクであり、softmax 関数を用いて各トラックである確率 (p) を最終層で出力する。例えば真のラベル (y) がタウトラックである場合 $y=[1,0,0,0]$ と表す。最終層の出力が $[0.7,0,1,0,1,0.1]$ であった場合、次のように計算する。

$$loss = -(1 * \log 0.7 + 0 * \log 0.1 + 0 * \log 0.1 + 0 * \log 0.1) \quad (5.2)$$

タウ粒子同定の損失とトラック識別の損失を合計し、逆伝播させることで重みを更新し、学習を進める。またトレーニングデータの真のタウトラック、対生成トラック、孤立トラック、フェイクトラックの数は偏りがあり、フェイクトラックが大半を占めている。よってそれぞれの真のトラックの数の逆数を重み付けすることで、偏りのない学習をしている。具体的には式 (5.2) の計算をする際に重みを掛け算している。

5.2 トラック識別の性能評価

トラック識別の評価方法について説明する。それぞれのトラックについてどの種類のトラックであるかの確率が $[0.6, 0.2, 0.1, 0.1]$ のよう出力されるので、一番数値が大きいカテゴリにそのトラックは分類される。表 5.1 の条件でカットした信号事象の真の prong 数が 1 本と 3 本の場合について、それぞれ評価を行った。

現行手法による結果を図 5.3 に示す。TT はタウトラック、CT は対生成トラック、IT は孤立トラック、FT はフェイクトラックのことである。真のトラックの種類が横軸で、RNN によって分類されたトラックの種類が縦軸になっている。それぞれの真のトラックの数を分母として、どのトラックに分類されたかの割合を書いている。例を挙げると、truth 1-prong の結果において真のタウトラックの内 95.5% が正しくタウトラックとして識別されている。GNN を用いた手法の結果を図 5.4 に示す。信号事象、背景事象ともに 1.6M ずつ用いて学習を行った結果である。現行手法と比べて、孤立トラックの識別が良くない。これは 5.1 項の学習で説明したトレーニングデータ内のそれぞれの真トラックの数で重みづけが影響しているかもしれない。重みづけを変更して、改善するか調査する必要がある。また全体のデータサンプル数が足りない可能性もあるので、データサンプル数を増やして改善するかも調査する必要がある。

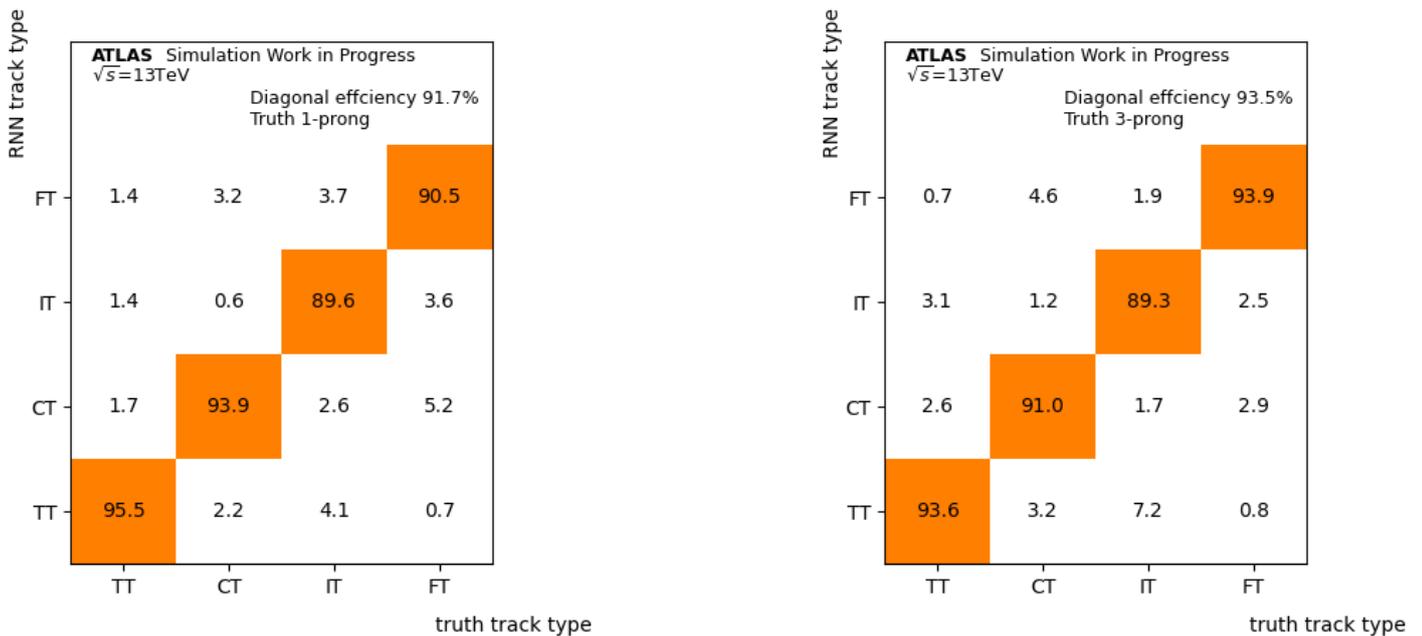


図 5.3: 現行手法の信号事象のトラック識別効率 (左: Truth 1-prong, 右: Truth 3-prong)

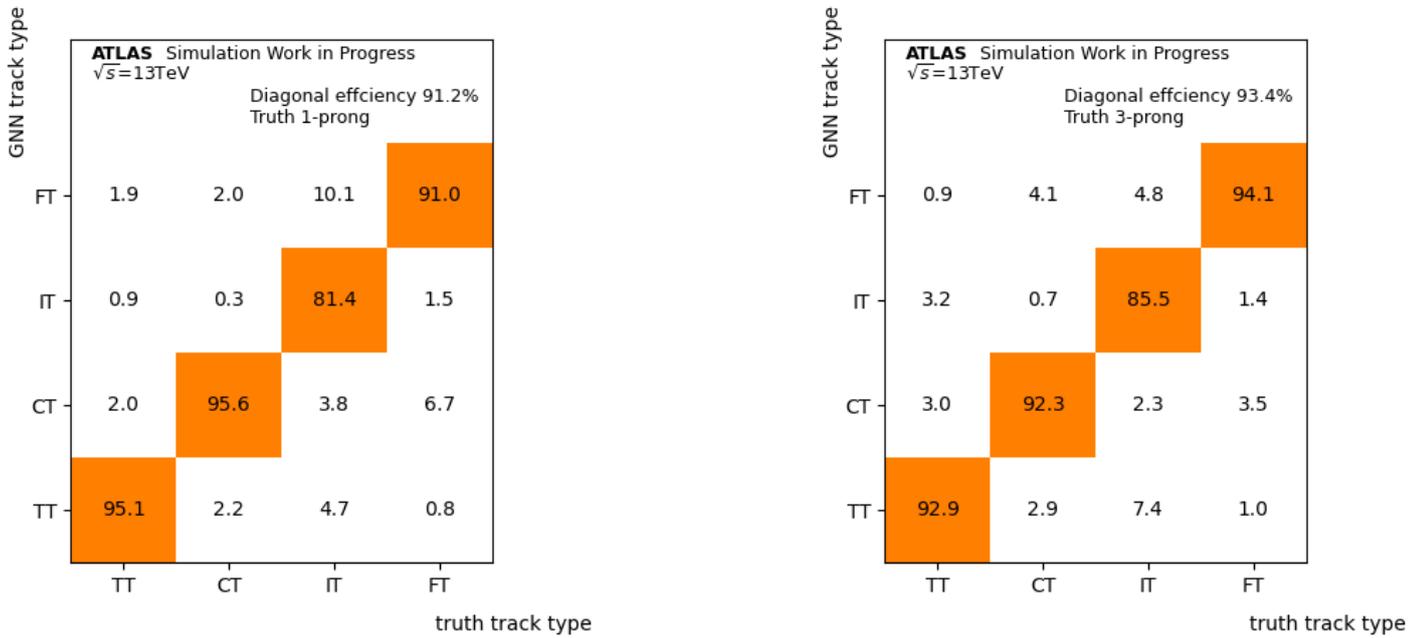


図 5.4: GNN を用いた手法の信号事象のトラック識別効率 (左: Truth 1-prong, 右: Truth 3-prong)

次に信号事象 ($p_T > 20 \text{ GeV}$) の truth 1-prong、truth 3-prong において、いくつかの prong として識別されたかを現行手法と GNN を用いた手法で比較した結果を図 5.5、図 5.6 に示す。横軸は識別された prong 数で縦軸は割合となっている。truth 1-prong かつ 1prong として識別された割合は現行手法と GNN を用いた手法で 1%以下の違いであるが、truth 3-prong かつ 3prong の割合は 2%ほど GNN を用いた手法の方が低い。

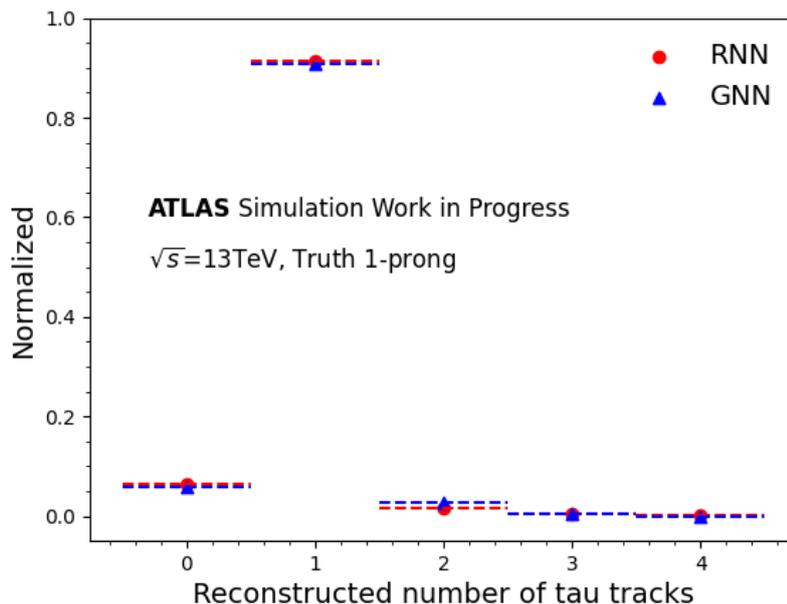


図 5.5: 現行手法と GNN を用いた手法における信号事象の prong 数 (Truth 1-prong)

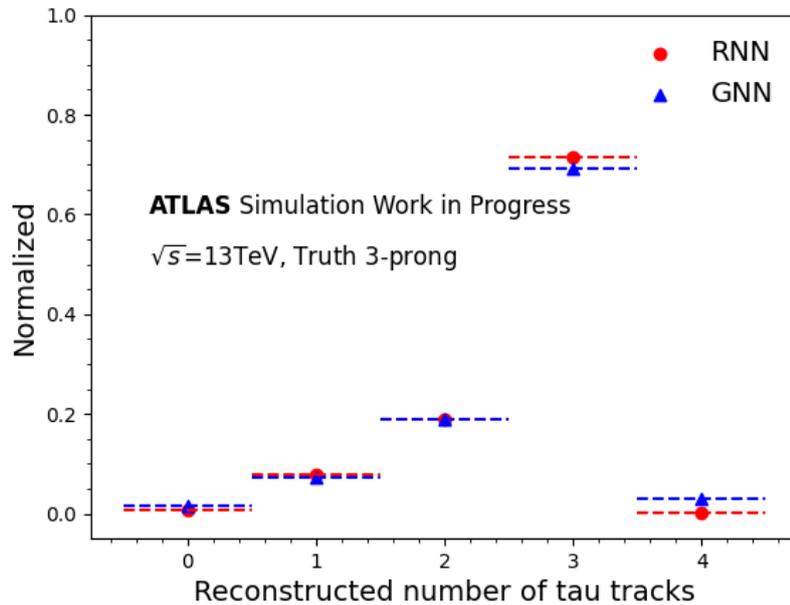


図 5.6: 現行手法と GNN を用いた手法における信号事象の prong 数 (Truth 3-prong)

また図 5.7 に背景事象 ($p_T > 20$ GeV) の prong 数の識別結果を示す。横軸は識別された prong 数で縦軸は割合となっている、prong 数が 5 より大きい場合は 4 のビンにオーバーフローしている。GNN の手法では現行手法と比べて、1prong が約 1.3 倍になっているのに対して、3prong は約 0.25 倍になっている。

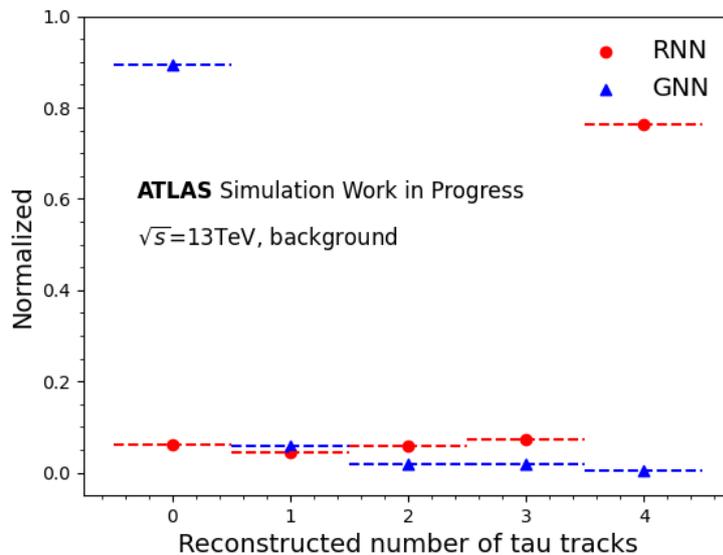


図 5.7: 現行手法と GNN を用いた手法における背景事象の prong 数

背景事象 300,000 サンプルのうち現行手法では 3prong として識別されるものが 22103 あり、GNN を用いた手法では 5,503 ある。現行手法で 3prong として識別された 22,103 サンプルは GNN を用いた手

法でいくつの prong 数と識別されたかの結果を図 5.8 に示す。22,103 のうち 15,722 は 0prong と識別され、3,507 は 3prong と識別された。

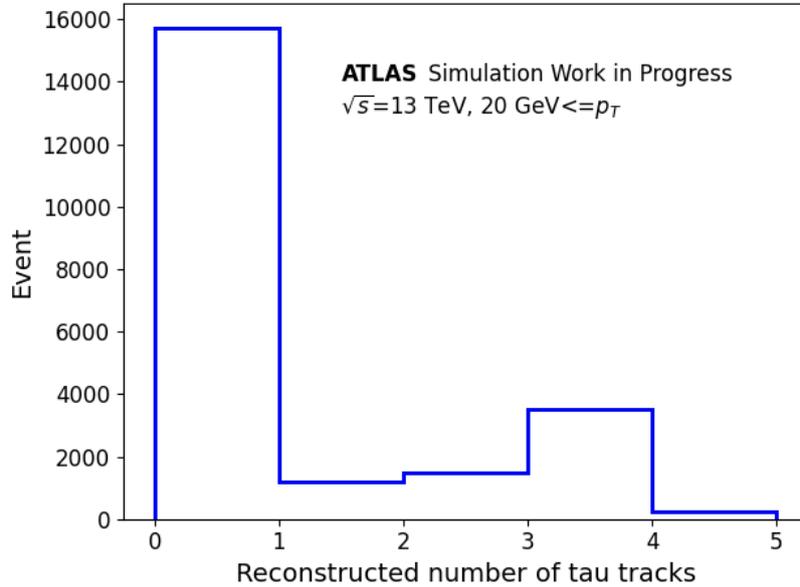


図 5.8: GNN を用いた手法によって識別された prong 数 (現行手法で 3prong と識別されたイベントのうち)

0prong と 3prong と識別されたものそれぞれの RNNscore 分布を図 5.9 に示す。図の緑は現行手法で 3prong で識別されたもので、青線はそれぞれ 0prong と 3prong として GNN に識別されたものである。現行手法で 3prong と識別かつ GNN の手法で 3prong と識別されたものは RNNscore が高く、背景事象と認識されずらいイベントということがわかる。逆に 0prong と識別されたものは RNNscore が低く、背景事象と認識されやすいイベントが多く、GNN の手法では 3prong と識別されていないことがわかる。

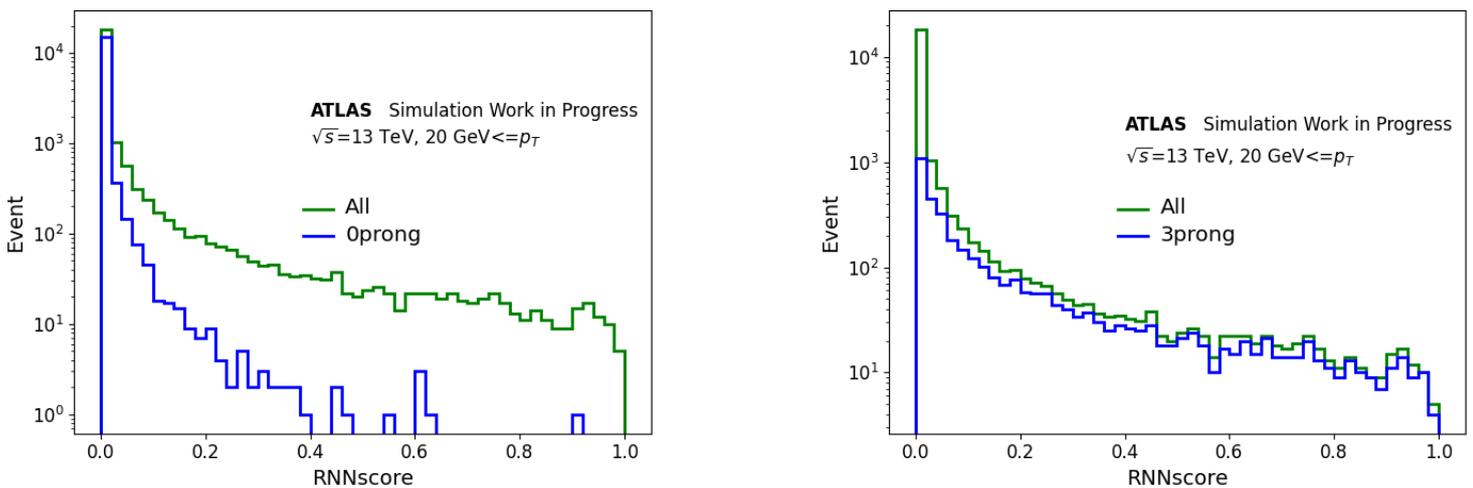


図 5.9: 0prong と 3prong と識別されたものの RNNscore 分布

5.3 同定の性能評価

次にタウ粒子同定の性能評価を行う。現行手法と比較するにあたって注意しなければいけないことがある。それは現行手法では RNN によりトラック識別が行われており、今回新たに行った GNN によるトラック識別と性能が異なるからである。それぞれの性能の違いについては 5.2 項で説明した。したがって信号事象の取得効率はスコア閾値より大きい事象/truth 1-prong、背景事象除去能力は背景事象 (すべての prong 数)/スコア閾値より大きい背景事象 (1prong) で計算する。3prong も同様である。このように prong 数が決まる前の事象を全体に設定することで、トラック識別の性能も加味した評価を行うことができる。現行手法 (赤) と 4 章で説明した GNN を用いた手法 (青) とトラック識別も含めた GNN の手法 (緑) の性能比較を図 5.10 に示す。トラック識別も含めた GNN の手法は 3.2M のデータサンプル数を用いたものである。1prong においては現行手法、トラック識別も含めた GNN の手法、4 章で説明した GNN を用いた手法の順番に性能が良い。3prong においては現行手法とトラック識別も含めた GNN の手法が同程度、4 章で説明した GNN を用いた手法の方が性能が良いという結果になった。

トラック識別も含めた GNN の手法の性能が 4 章で説明した GNN を用いた手法よりも悪くなってしまった原因はいくつか考えられる。まず一つは信号事象の取得効率である。図 5.4、図 5.5 に示すように truth 1-prong が 1-prong と正しく識別できる割合が現行手法よりも悪くなっている。(3prong も同様) データサンプル数を増やす、トラック識別における重み付けの最適化を行うことで、信号事象の取得効率が上がり、同定の性能も向上すると考えられる。もう一つは学習におけるデータサンプル数が少ないという点である。4 章で説明した GNN を用いた手法は 1prong、3prong のそれぞれに特化した学習である。それに対して、トラック識別も含めた GNN の手法では信号事象に関して、truth 1-prong、truth 3-prong の両方を用いて学習を行っており、データサンプル数が 4 章で説明した GNN を用いた手法よりも少ない。また背景事象においても図 5.6 に示したように 1prong、3prong と識別されるものは少なく、背景事象の学習が十分に行えていない可能性がある。データサンプル数を増やす、学習における重み付けを最適化することで今後改善を目指す。

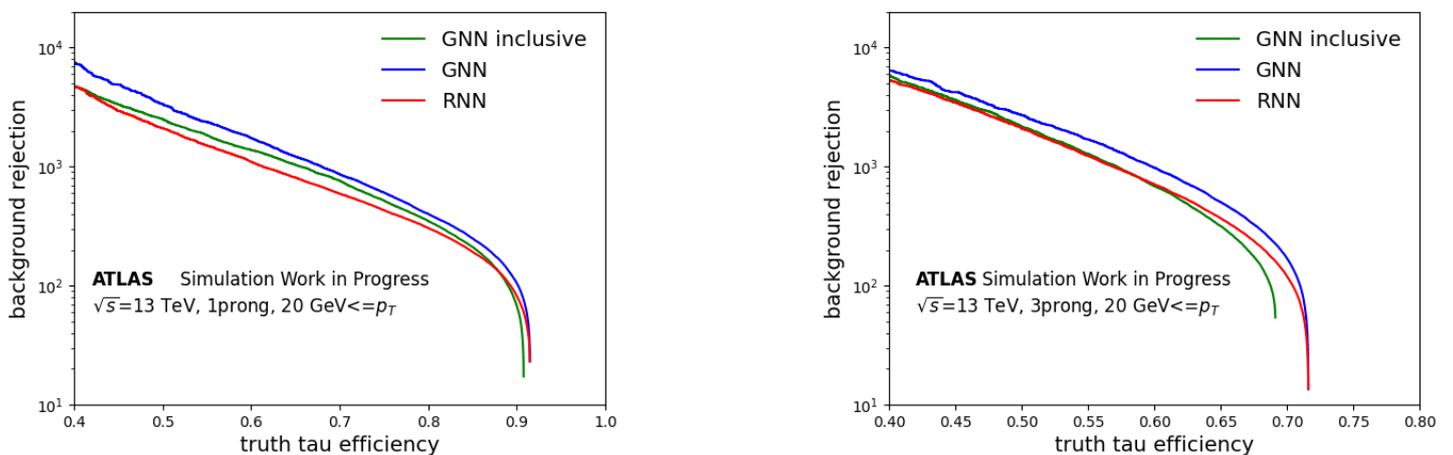


図 5.10: 現行手法との性能比較 (左: 1prong、右: 3prong)

6 結論と今後の展望

結論

ハドロニック崩壊するタウ粒子は QCD ジェットとの区別が難しいため、機械学習による同定が行われている。タウ粒子を用いた解析にはヒッグス粒子の精密測定、重いヒッグス粒子探索、スタウ探索など様々あり、低横運動量から高横運動量のタウ粒子まで必要とされている。これらのタウ粒子を用いた解析の感度向上のために、同定の性能向上が必要である。本研究では新たにグラフニューラルネットワーク (GNN) を用いたタウ粒子同定手法の開発を行った。

パラメータ調整として GNN の層数と次元の調整を行い、層数を増やす方が性能向上に効果的であることがわかった。またデータサンプル数を増やすことで、学習モデルのパラメータ数も大きくすることができ、同定の性能が向上することを確認できた。1prong では 3.2M、3prong では 1.6M のデータサンプルを用いることで現行手法よりも 1.3 から 1.4 倍程度背景事象除去能力が向上した。

またよりタウ粒子同定の性能を向上させるために新しい学習モデルを検討した。タウ粒子同定の前に行われているトラック識別も同時に行うことで、トラック識別による学習がタウ粒子同定の性能向上をもたらす可能性を検討した。信号事象の prong 数の識別においては現行手法と同程度の性能であったが、背景事象の prong 数の識別は現行手法と GNN を用いた手法で性能が異なった。これらのトラック識別の性能の違いを考慮し、現行手法と GNN を用いた手法の同定性能の比較を行ったところ、著しい向上は見られなかった。

原因はいくつか考えられる。1つ目は信号取得効率が現行手法より悪いことである。データサンプル数を増やす、トラック識別における重み付けの最適化を行うことで、信号事象の取得効率が上がり、同定の性能も向上すると考えられる。2つ目は図 5.7 に示すように 1prong,3prong と識別される背景事象の割合が少ないことである。同定においては 1prong,3prong 以外のイベントはノイズとなる可能性があり、十分背景事象の学習が行えていないことが考えられる。現行手法ではあらかじめ 1prong,3prong と識別されたイベントを学習に使用するので上記のようなことは起こらない。これらの原因を解決するためには、データサンプル数を増やすことがまず考えられる。データサンプル数を増やせば、1prong,3prong と識別される背景事象の絶対数が増えるためである。

今後の展望

トラック識別を含めたタウ粒子同定手法の改善をデータサンプル数を増やす、重み付けの最適化などによって行っていく。

また本研究は MC を用いた性能評価であるため、今後は実データに使用できるか検証する必要がある。機械学習に用いている入力変数分布が実データと MC で一致しているか確かめ、実データでも同様の性能が発揮されるか検証する。

また図 5.5、図 5.6 に示すように truth 1-prong、truth 3-prong のタウ粒子が 2prong として識別される場合がある。この 2prong のタウ粒子を解析に用いることで信号の統計量が増え、解析の感度を向上することができる。とくに truth 3-prong が 2prong に識別される割合が大きく、図 6.1 に示すように高横運動量領域でより大きくなる。また、2prong は再構成される電荷が $-2, 0, +2$ である。タウ粒子の電荷を用いる解析のために、2prong の電荷を正しく識別する必要がある。したがって 2prong のタウ粒子の電荷識別と同定への応用可能性についても検討することが重要である。

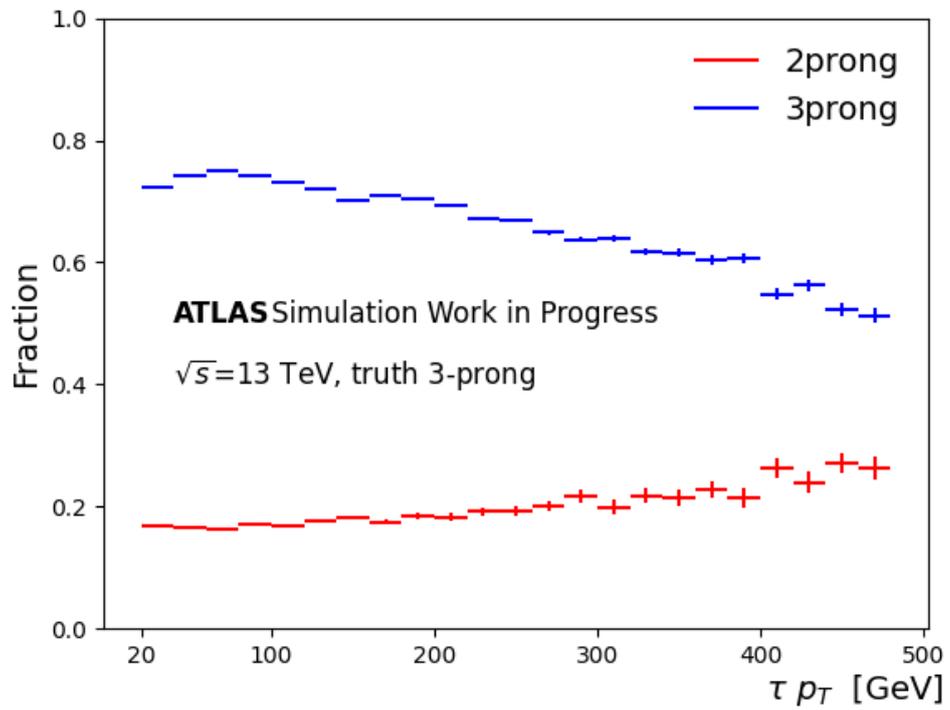


図 6.1: 横運動量領域における信号事象 truth 3-prong の prong 数

7 謝辞

まず初めに、3年間研究指導をしてくださった寄田浩平教授に感謝申し上げます。機械学習を用いた研究を行いたいという私の希望も考慮して、グラフニューラルネットワークを用いたこの研究テーマを与えていただきありがとうございました。毎週のミーティングでは最終結果を見せるだけでなく、その結果に至った過程も説明しなさいなど様々なアドバイスをいただきました。これらの研究に対する姿勢をこれからも活かしていきたいと思えます。ありがとうございます。

次に、田中雅士主席研究員にも感謝申し上げます。中間発表などの場において、研究内容に対する指摘・アドバイスをいただきました。私自身が考えに至っていない部分を教えていただきました。ありがとうございます。

坂本敦子秘書には事務の手続きで大変お世話になりました。また定期的で開催していただいた卓球の集まりは研究の息抜きになりました。ありがとうございます。

ATLAS 実験の先輩である秋山大也さん・伊藤紘貴さんには大変お世話になりました。秋山さんには毎週のミーティングで、こういうプロットも見た方がいいのではなどの研究方針の指摘・アドバイスをたくさんいただきました。伊藤さんにはタウ粒子の研究に関連していることもあり、ソフトウェアの使い方を教えていただいたり、研究でわからないことについて聞かせていただきました。同期の清水虎冨さん、谷口日奈子さんは研究グループは違いましたが、GRAMS 実験で頑張っている姿を見て刺激をもらいました。ありがとうございます。

また ATLAS 日本の Working Group 1 のミーティングメンバーにも感謝申し上げます。毎週のミーティングで研究に対するアドバイス・議論等をしていただきました。皆様のご協力なくして修士論文を完成させることはできませんでした。ありがとうございます。とくに東大素粒子センターの森永真央さんには大変お世話になりました。この研究を進めるにあたり、機械学習周りの技術的な部分をたくさん教えていただき、研究の方針のアドバイスもしていただきました。ありがとうございます。

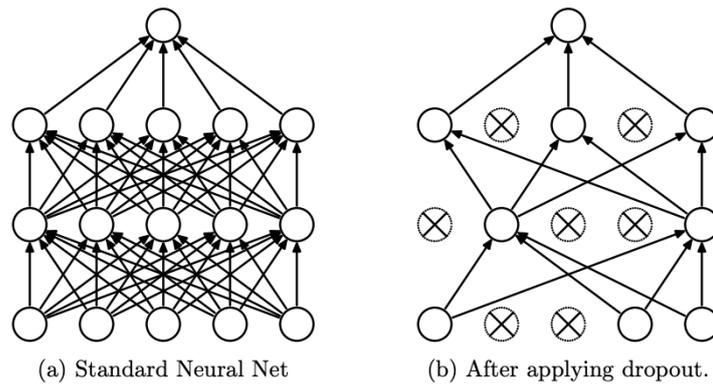
最後に、両親には生活面で大変お世話になりました。大学のみならず、大学院まで通わせてもらえたことに感謝しています。本当にありがとうございます。

A 学習におけるパラメータ調整

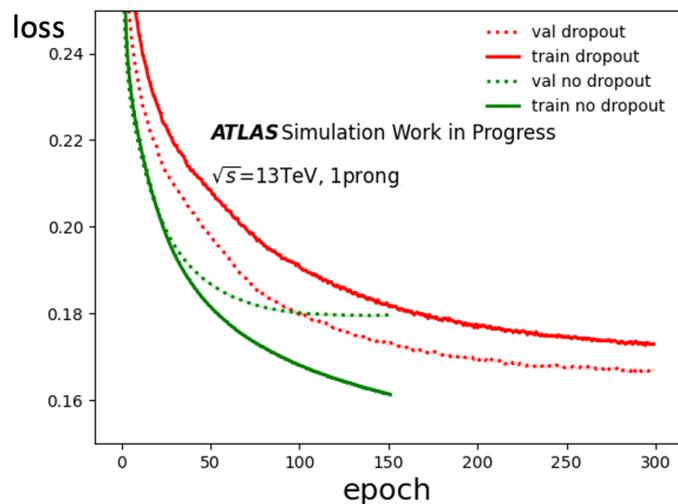
GNN を用いた機械学習を行う際に、過学習が起きないように幾つかのパラメータを調整して行ったのでそれについて説明する。過学習とはトレーニングデータに学習モデルが順応しすぎて、未知のデータに対しては性能が下がってしまうことである。

Dropout

Dropout とは特定のレイヤーの出力を学習時に一定の確率でランダムに 0 にする手法であり、イメージ図を図付録 1 に示す。一部のデータが欠損していても正しく認識でき、局所的な特徴が過剰に評価されるのを防ぐ。図付録 2 に GNN を用いたタウ粒子同定において Dropout を使用したとき (赤) としなかったとき (緑) の学習時の損失の推移を示す。緑を見ると 50 エポック (学習回数) あたりからトレーニングデータ (実線) の損失が減少しているが、バリデーションデータ (点線) の損失は減少していない。これは過学習である。Dropout を使用した赤では、実線・点線ともに損失が減少しており過学習が起きていないことが分かる。



図付録.1: Dropout のイメージ図 [18]



図付録.2: Dropout の有無による学習の推移の違い (赤: Dropout あり、緑: Dropout なし)

B 転移学習

本章では学習におけるトレーニングデータのデータサンプル数を大きくすることで、モデルのパラメータ数も大きくすることができ、性能が向上することを説明した。しかし、シミュレーションの作成には膨大な計算資源を消費し、時間がかかるという問題がある。したがって、データサンプル数を多く用意できないときの性能の向上方法について説明する。

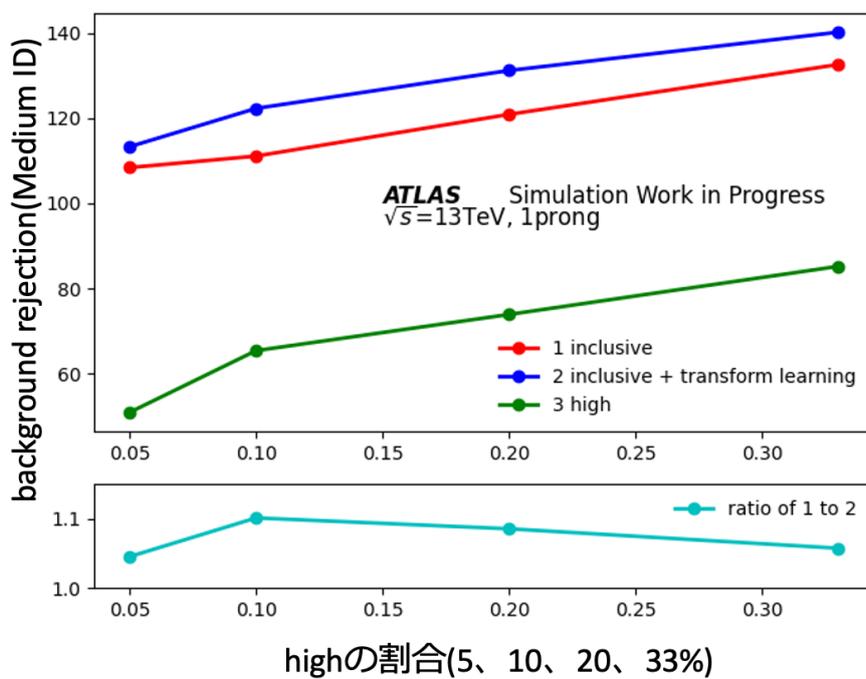
転移学習という機械学習で用いられている手法を使う。転移学習とはあるタスクで学習されたモデルを別のタスクに利用する手法である。例を挙げると猫の種類判別を行いたいときに、猫の画像データは少ないが犬の画像が多い場合を考える。犬の種類識別を行い、その学習モデルを再利用して猫の種類識別を行う。犬の種類判別によって得られた知識を、猫の種類判別モデルにも適応させることができる。

転移学習をタウ粒子同定手法にどう適用するか説明する。4.2 項でも説明しているが使用している MC はタウ粒子の横運動量の低いサンプルが多く、高いサンプルが少ない。したがって、イベントごとに重み付けをして汎用的な学習を行っていた。しかし、重みづけによる学習は複雑であり、性能の不安定化をもたらしてしまうデメリットがある。そこで一度重みづけをしないで学習を行ったあとに、特化させたい横運動量領域のサンプルで再学習することで上記の問題を解決することができる。今回はとくにデータサンプル数が少ない $p_T > 100$ GeV のサンプル (high とこれから呼ぶ) で再学習させ、そのデータサンプル数の割合を変化させた時の性能変化について調べる。次の 3 つの学習方法で比較する。

1. 全横運動量領域で学習 (赤)
2. 全横運動量領域で学習、high 領域で再学習 (青)
3. high 領域のみで学習 (緑)

全横運動量領域のデータサンプル数に対する high の割合を 5,10,20,33 % と変化させた時の結果を図付録 3 に示す。図の緑と赤を比較すると、赤の方が性能が良い。赤が全横運動量領域のサンプルを用いているため、high 領域以外のサンプルによる統計量のアドバンテージがあるからである。つまり、低横運動量領域のサンプルも high 領域の学習に寄与している。赤と青の比をとったものが水色で書かれていて、high の割合が 10 % のとき約 10 % のゲインがあり、high の割合が 10 % 以上になるとゲインが下がっている。これは high の割合が 5 % のときは転移学習に用いるデータサンプル数としては少ないが、10 % を超えてくると転移学習前の学習で十分に high 領域の学習ができているためゲインが減っていると考えられる。

このようにタウ粒子同定においてもデータサンプル数が少ない領域に特化した学習を行いたい場合、転移学習を用いることで性能の向上が見られる。



図付録.3: high の割合を変えた時の性能変化

参考文献

- [1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", <https://arxiv.org/abs/1207.7214>
- [2] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector", <https://arxiv.org/pdf/1610.07922.pdf>
- [3] ATLAS Collaboration, "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector", <https://arxiv.org/abs/2007.07830>
- [4] CMS Collaboration, "Evidence for Higgs boson decay to a pair of muons", <https://arxiv.org/abs/2009.04363>
- [5] ATLAS Collaboration, "A combination of measurements of Higgs boson production and decay using up to 139 fb⁻¹ of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment", <https://cds.cern.ch/record/2725733>
- [6] 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>
- [7] 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/abs/2002.12223>
- [8] 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/abs/1911.06660>
- [9] http://alice-j.sakura.ne.jp/wp-alice-j/wp-content/uploads/2018/02/CERN_Accelerator_Complex.png
- [10] The ATLAS Experiment at the CERN Large Hadron Collider, <https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003/pdf>
- [11] "The anti- k_t jet clustering algorithm", <https://arxiv.org/abs/0802.1189>
- [12] ATLAS Collaboration, "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1", <https://arxiv.org/abs/1603.02934>
- [13] 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/abs/1412.7086>
- [14] 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://arxiv.org/abs/1306.4945>
- [15] 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>
- [16] Graph Neural Networks: A Review of Methods and Applications, <https://arxiv.org/abs/1812.08434>
- [17] Inductive Representation Learning on Large Graphs, <https://arxiv.org/abs/1706.02216>

- [18] Dropout: A Simple Way to Prevent Neural Networks from Overfitting,
<https://jmlr.org/papers/v15/srivastava14a.html>