深層学習による軌道材料画像のレール目視検査モデルの開発

東日本旅客鉄道　正会員　○吉田　尚
東日本旅客鉄道　正会員　葛西　亮平
東北大学，理化学研究所革新知能統合研究センター　岡谷　貴之

1．はじめに

JR 東日本では，線路設備モニタリング装置を実用化し，2020 年度までに全 50 線区に導入が完了し，線路総合巡視等の効率化を実現した．同装置により取得されるデータは，それ以外にも様々な用途で活用が期待できる．そこで，今回はレール目視検査のために開発した深層学習モデルについて説明する．

2．課題設定

まず，現状のレール目視検査で管理している不良項目について，検査結果からその発生状況を確認した．結果は**図１**に示すとおりであり，シェリング，きしみ割れ，水平裂が全体の 88%を占めることが分かった．シェリングと水平裂は外観上の差はない（超音波探傷をしなければ区別できない）ため同一とみなし，今回は，これに波状摩耗を加えてシェリング，きしみ割れ，波状摩耗の 3 項目を検査するモデルを開発する事にした．

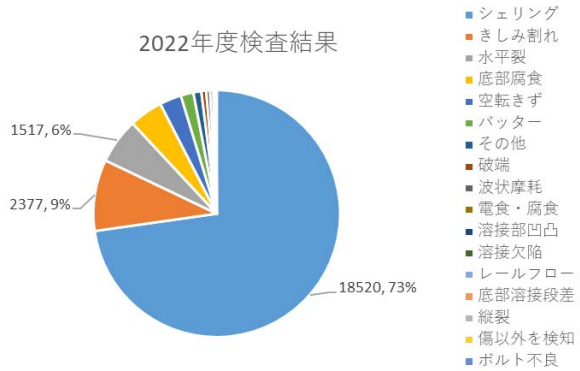


図１　モニタリング導入線区のレール目視検査結果

画像から不良箇所を検知する方法については，いくつか方法があるが，今回は以下の観点を重視して方法を選択した．

・レール全長を検査する必要があるため，計算コストが小さいこと
・アノテーションの労力が小さいこと
・同一箇所に複数の不良が混在する状況に対応できること（例：シェリング＋きしみ割れ）

上記を考慮の上，**図２**のモデルとした．Backbone には軽量な MobileNetV2[1)] をベースとし，パラメータを削減する工夫をして全体で 553,854 のパラメータに抑えた．出力は 6 つのロジットであり，シグモイド関数で確率化して順序回帰[2)] により画像を分類する．なお，入力画像を作成する前処理として，画像全体からレールを中心に高さ 512，幅 96 を切り出す．

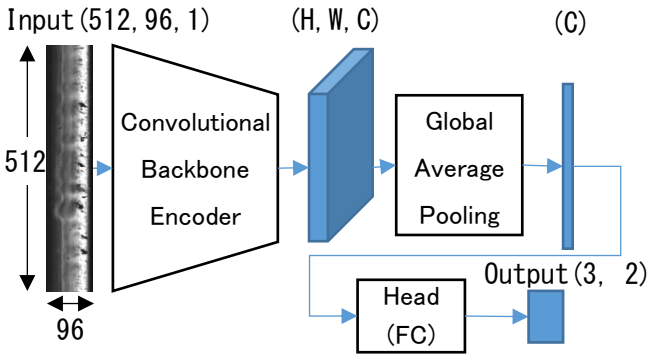


図２　開発したモデル構造

出力は，3×2 の配列であり，縦軸は分類項目（シェリング，きしみ割れ，波状摩耗），横軸は**表１**に示すような順序の確率を意味している．このような順序を設定した理由は以下による．

・不良の程度をある程度予測したい
・明確な判断が難しいサンプルにおけるアノテーションの労力軽減及びモデル学習時におけるサンプルのバイアスの影響軽減

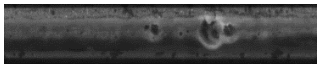
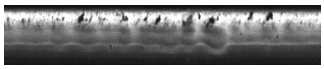
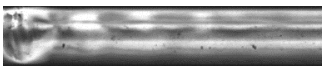
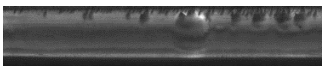
表１　出力の内容

	0:Rank0	1:Rank1
0:シェリング	疑わしい	シェリング有
1:きしみ割れ	きしみ割れ	GC き裂
2:波状摩耗	疑わしい	波状摩耗有

キーワード　軌道材料画像，深層学習，レール目視検査
連絡先　〒331-8513 埼玉県さいたま市北区日進町 2－479
JR 東日本研究開発センター　テクニカルセンター　TEL048-651-2389

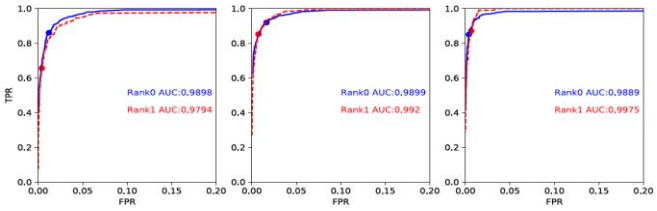
サンプル画像とラベルの例は**表２**に示すとおりであり，1つの画像に対して，3つの判定項目について，正常を含めてそれぞれ3段階で状態を定義した．

表２　サンプル画像とラベルの例

画像	ラベル	意味
	[[1, 1], [0, 0], [0, 0]]	シェリング有
	[[1, 0], [1, 0], [0, 0]]	シェリング疑わしい， きしみ割れ
	[[1, 1], [0, 0], [1, 0]]	シェリング有， 波状摩耗疑わしい
	[[1, 0], [1, 1], [0, 0]]	シェリング疑わしい， GC き裂

３．結果

サンプルは29,988枚用意し，うち20,991枚を訓練データ，8,997枚を検証データとしてモデルを学習した．検証データに対するROC曲線は**図３**に示すとおりであり，非常に高い精度を実現することができた．なお，閾値が0.5の場合を点でプロットしている．



a) シェリング　　b) きしみ割れ　　c) 波状摩耗

図３　検証データに対するROC曲線

上記の結果は検証データに関するものであり，モデルの精度を評価するためには，別途テスト用のデータが必要である．また，上記のサンプルは1人でアノテーションしたものであるが，モデルの客観性を確認するためには複数人で結果を確認したほうが良い．そこで，所内の15人に一人最低1,000枚ずつ分散してアノテーションを依頼し，改めて精度を評価した．確認する画像は，367,416枚の画像に対して推論をかけて，いずれかの不良が判定された66,000枚からランダムに割り当てた．

15人による確認データのサンプルサイズを揃えるために，確認済みの画像を確認者1人に対して1,000枚リサンプリングし，計15,000枚の画像に対して精度を確認した結果を**表３**に示す．サンプルの中には画質が悪いものがあり，人間では判定できないものは除外したため，サンプルの合計は15,000に一致しない．AIの予測結果は，閾値を0.5で分類したものである．結果を見ると，いずれの項目もRank0よりもRank1の精度が低くなっている．きしみ割れのRank1のPrecisionが低いが，これはAIがGCき裂と予測したサンプルの中に人間がきしみ割れと判断した画像が多く含まれていたためである．そのようなサンプルの例を**図４**に示すが，きしみ割れ区間にシェリングのような落ち込みが併発しており，GCき裂があるようにも見える画像が多く見られた．このように，人でも判断に迷うような画像が精度に現れていると考えられる．なお，実運用場面を考えれば，Rank1＝1のデータが，Rank0で検出されていれば大きな見逃しは無いと考えられ，そのような見逃しの確率は，閾値＝0.5で現状2%前後となっている．

表３　AI予測結果の人間による目視確認結果

シェリング		AI予測 正常	疑わしい	傷あり	計
人間	正常	8,493	534	23	9,050
	疑わしい	506	3,625	137	4,268
	傷あり	27	372	1,233	1,632
計		9,026	4,531	1,393	14,950

	Precision	Recall
Rank0	90.6%	91.0%
Rank1	88.5%	75.6%

大きな見逃し
1.7% （＝27／1632）

きしみ割れ		AI予測 正常	きしみ割れ	GCき裂	計
人間	正常	3,619	453	81	4,153
	きしみ割れ	207	7,671	1,104	8,982
	GCき裂	38	278	1,499	1,815
計		3,864	8,402	2,684	14,950

	Precision	Recall
Rank0	95.2%	97.7%
Rank1	55.8%	82.6%

大きな見逃し
2.1% （＝38／1815）

波状摩耗		AI予測 正常	疑わしい	波状摩耗有	計
人間	正常	13,873	124	13	14,010
	疑わしい	70	537	69	676
	波状摩耗有	7	36	221	264
計		13,950	697	303	14,950

	Precision	Recall
Rank0	86.3%	91.8%
Rank1	72.9%	83.7%

大きな見逃し
2.7% （＝7／264）

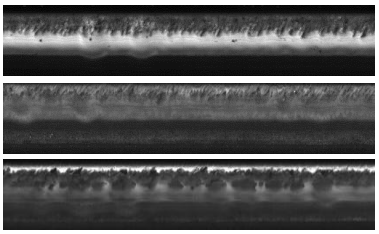


図４　AI予測GCき裂→人間きしみ割れ判定の画像

４．おわりに

深層学習による軽量でマルチタスクに対応したレール目視検査モデルを開発する事ができた．今後は実用化に向けて，最終的な調整を進めていく．

参考文献

１）Mark Sandler et al.: MobileNetV2: Inverted Residuals and Linear Bottlenecks, IEEE Conference on Computer Vision and Pattern Recognition, 2018
２）Zhenxing Niu et al.: Ordinal Regression with Multiple Output CNN for Age Estimation, IEEE Conference on Computer Vision and Pattern Recognition, 2016