

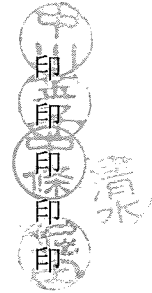
(様式 11)

令和 2 年 2 月 1 0 日

学 位 論 文 審 査 要 旨 (課程博士)

東京農工大学大学院工学府長 殿

審査委員 主査 中川 正樹
副査 金子 敬一
副査 中條 拓伯
副査 清水 郁子
副査 堀田 政二



学位申請者	電子情報工学専攻 2017 年度入学 学籍番号 17834303
	氏 名 NGUYEN CONG KHA
申請学位	博 士 (工学)
論文題目	中国語起源の歴史文書認識の試み
論文審査要旨 (600~700 文字) 本論文は、中国語起源の歴史的な文書認識、特にベトナムの古語であるチュノム文書と日本の木簡画像の処理と認識を対象にしている。歴史的な文書認識は今、多くの国で取り組まれている。なぜなら、その時代の文化、経済、政治について重要な価値を持っているからである。歴史的な文書を認識するためには、まず高精度の OCR (光学文字認識) が必要である。本論文では、著者は、ノイズ除去、文字分割、認識処理、後処理などの、OCR を作成するためのさまざまな手法を提示している。各アプローチのメリットとデメリットを実験で確認し、具体的なシステムを構築した。ベトナムで 13 世紀から 20 世紀まで公用語として使われたチュノム文書の解読と電子化では、基本システムと深層ニューラルネットワークによる更新版の 2 つシステムを開発した。基本システムでは前処理、文字分割、文字認識、ユーザによる結果修正の 4 つの部分からなる。更新システムでは畳み込みニューラルネットワークでそれぞれの部分を更新した。また、言語文脈を組み合わせ、基本システムの認識率 (69.08%) より 16 ポイント以上改善した 85.22% の認識率を達成した。奈良平城京から出土する木簡画像の復元では、畳み込みニューラルネットワークである Generative Adversarial Network の一形態を採用した。この手法は二値化と同じ効果をもたらし、従来の二値化に勝る結果を達成した。一部が欠けた文字、	

(様式 11)

かすれた文字も復元できるようになった。処理されたパターンの復元程度を評価するために、学習された光学的読取り方式（OCR）を適用し、伝統的な二値化方法で処理されたパターンに比べて、提案方法で復元されたパターンで、最も高いOCR認識率が得られた。

審査経過（時系列）

2019年12月11日 令和2年3月博士後期課程修了に係る学位申請

2019年1月15日 審査委員の選出・指名・付議，論文審査委員の付託（運営委員会）

2020年1月30日 学位論文発表会（本審査）

2019年2月10日 本専攻内情報工学分野における博士学位取得要件「学術論文誌掲載論文1件，査読付き国際会議発表論文1件」（WoS論文1報を含む）を満たしていることを確認の上，専攻会議で論文合格及び最終試験合格を承認。

2019年3月3日 学位授与認定・修了認定（運営委員会）